Perfect matchings, Hamilton cycles, degree distribution and local clustering in Hyperbolic Random Graphs

Schepers, Markus

DOI:
10.33612/diss.124993623

Perfect matchings, Hamilton cycles, degree distribution and local clustering in Hyperbolic Random Graphs

Markus Schepers

PhD thesis, University of Groningen, The Netherlands

Printed by: Ridderprint, The Netherlands
Cover: Galaxy picture by Graham Holtshausen,
see https://www.grayphotography.com.au/,
simulation of KPKVB model by the author

ISBN: 978-94-034-2672-3 (printed version)
ISBN: 978-94-034-2673-0 (electronic version)

rijksuniversiteit
groningen

Perfect matchings, Hamilton cycles, degree distribution and local clustering in Hyperbolic Random Graphs

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 22 mei 2020 om 14.30 uur

door

Markus Schepers

geboren op 16 juli 1993
te Freiburg im Breisgau, Duitsland

Promotores
Prof. dr. T. Müller
Prof. dr. E. C. Wit

Beoordelingscommisie
Prof. dr. R. van der Hofstad
Prof. dr. K. Panagiotou
Prof. dr. D. Valesin

Contents

Chapter 1

Introduction

In this thesis, we will study several properties of a model of random graphs that involves points taken randomly in the hyperbolic plane. Random graphs are a mathematical model for networks, i.e. systems which consist of several entities, e.g. points, people or web sites, and pairwise relationships between these entities, e.g. line segments, friendships or hyperlinks. In graph theory, the entities are called vertices and the relationships between them are called edges. In networks science, the vertices are also often called nodes and the edges are often called links.

In this introductory chapter, we will firstly give a brief motivation for hyperbolic geometry and define the random graph model that we want to study. Then, in Sections 1.3, 1.4 and 1.5, we will introduce some graph-theoretical concepts and present the corresponding results that form the novel contributions of this thesis and that we will prove in the main part. We conclude the introductory chapter with an overview of related models and tools which constitute crucial proof ideas. The remaining chapters contain the detailed proofs.

1.1 Hyperbolic geometry

Hyperbolic geometry was developed in the first half of the 19th century in order to show that Euclid's fifth axiom was indeed independent of the others, i.e. that it is possible to have a line l and a point P not on it such that there are infinitely many lines through P that are parallel to l, while still all other axioms of Euclid hold (for a modern English translation of Euclid's Elements, including the Greek original in a column to the left-hand side, see [16]). The first ground-breaking publications on hyperbolic geometry were by Lobachevsky in 1829-30 (for the first English translation in 1891, see [43]) and independently by Bolyai in 1832, see [14], while Gauss mentioned some results in a letter from 1824, which he did not publish [15]. Later, hyperbolic geometry has been constructed and studied with analytic methods. This led to the modern description of the hyperbolic plane $\mathbb{H}$ as a surface with constant negative Gaussian curvature. It has several convenient representations (i.e. coordinate maps), including the Poincaré half-plane model, the Poincaré disk model and the Klein disk model. A gentle introduction to hyperbolic geometry and these representations of the hyperbolic plane can for instance be found in [48]. Throughout this thesis we will be working with a representation of the hyperbolic plane using *hyperbolic polar coordinates*. That is, a point $p \in \mathbb{H}$ is represented as (r, θ), where $r = r(p)$ is the hyperbolic distance between p and the origin (by which we mean a distinguished point O) and θ is the angle between the line segment Op and the positive x-axis. We shall denote by $\mathcal{D}_R$ the hyperbolic disk of radius R around the origin O of the hyperbolic plane $\mathbb{H}$ with curvature -1, and by $d_{\mathbb{H}}(u, v)$ we denote the hyperbolic distance between two points $u, v \in \mathbb{H}$. In polar coordinates, the hyperbolic distance between $u_1 = (r_1, \theta_1)$ and $u_2 = (r_2, \theta_2)$ can be computed explicitly via

$$d_{\mathbb{H}}((r_1, \theta_1), (r_2, \theta_2)) = \operatorname{acosh}(\cosh r_1 \cosh r_2 - \sinh r_1 \sinh r_2 \cos(\theta_2 - \theta_1)),$$

where $\operatorname{acosh} : [1, \infty) \to [0, \infty)$ denotes the inverse of the hyperbolic cosine function $\cosh : [0, \infty) \to [1, \infty)$, $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ and where we recall that $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$. Note that the expression for $d_{\mathbb{H}}$ is well-defined because the argument inside acosh is $\geq \cosh r_1 \cosh r_2 - \sinh r_1 \sinh r_2 = \cosh(r_1 - r_2) \geq 1$.

As we mentioned earlier that the hyperbolic plane is a surface with constant negative Gaussian curvature, we recall that curvature is a widely applicable concept in geometry. The curvature at a point of a (2-dimensional) surface measures by how much the surface deviates from a plane (close to that point). Zero curvature indicates that the surface locally looks like the plane. Positive curvature indicates that it locally looks like a sphere or tennis ball, or in other words, that the surface lies entirely on one side of the tangent plane. Negative curvature indicates that it locally looks like a hyperboloid or saddle for horseback riding, or in other words, the surface lies on both sides of the tangent plane. Roughly speaking, the larger the absolute value of the curvature, the more sharply bent or curly the surface. As we do not need it for our purposes, we do not develop the full theory of curvature here, but we refer to standard textbooks on differential geometry, for instance do Carmo [22].

1.2 The KPKVB model

The model of random graphs that we study in this thesis was introduced by Krioukov, Papadopoulos, Kitsak, Vahdat and Boguñá [36] in 2010 - we abbreviate it as *the KPKVB model*. We should however note that the model also goes by several other names in the literature, including *hyperbolic random (geometric) graphs* and *random hyperbolic graphs*. The model was intended to model complex networks and, in particular, it is motivated by the assumption that the properties of complex networks are the expression of a hidden geometry which expresses the hierarchies among classes of nodes of the network. Krioukov et al. postulate that this hidden geometry is hyperbolic space.

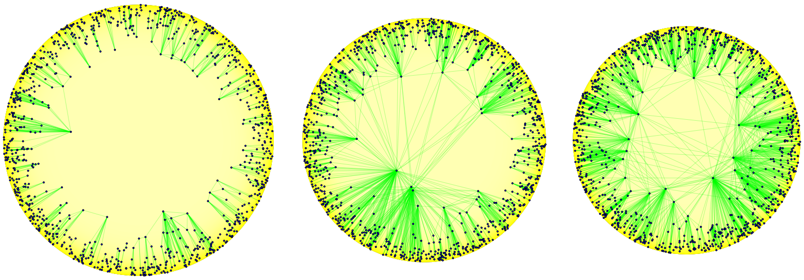

Figure 1.1: Simulations of the KPKVB model with $n = 1000$ vertices, $\alpha = 0.9$ and $\nu = 1, 2, 3$ (from left to right).

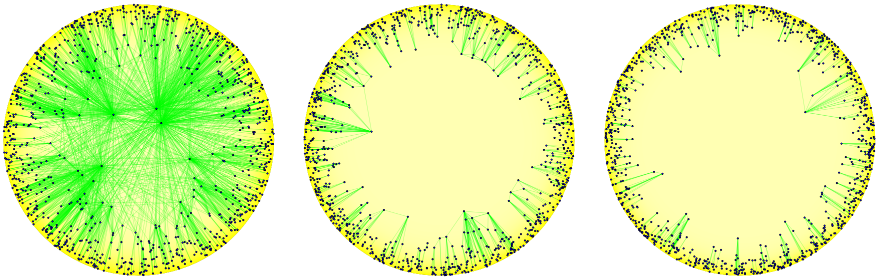

Figure 1.2: Simulations of the KPKVB model with $n = 1000$ vertices, $\nu = 1$ and $\alpha = 0.6, 0.9, 1.2$ (from left to right).

Given a fixed constant $\nu \in (0, \infty)$ and a natural number $n > \nu$, we let $R = 2\log(n/\nu)$, or equivalently $n = \nu \exp(R/2)$. Also, let $\alpha \in (0, \infty)$.

The vertex set of the KPKVB random graph $G(n; \alpha, \nu)$ consists of n i.i.d. points in $\mathcal{D}_R$ with probability density function

$$g(r, \theta) = g_{\alpha,R}(r, \theta) = \frac{\alpha \sinh \alpha r}{2\pi(\cosh \alpha R - 1)}, \tag{1.1}$$

for $0 \leq r < R$ and $-\pi < \theta \leq \pi$. The edge set is given by

$$\begin{aligned}
E &= \{uv \in \binom{V}{2} : d_{\mathbb{H}}(u, v) \leq R\} \\
&= \{(r_u, \theta_u), (r_v, \theta_v) \in \binom{V}{2} : \\
&\qquad\qquad \cosh r_u \cosh r_v - \sinh r_u \sinh r_v \cos |\theta_u - \theta_v|_{2\pi} \leq \cosh R\} \\
&= \{(r_u, \theta_u), (r_v, \theta_v) \in \binom{V}{2} : |\theta_u - \theta_v|_{2\pi} \leq \vartheta(r_u, r_v)\},
\end{aligned}$$

where $|x|_m = \min(|x|, m - |x|)$ for $m > 0$, $x \in \mathbb{R}, -m \leq x \leq m$, and where we used the angle threshold function $\vartheta = \vartheta_R = \vartheta_{n,\nu}$:

$$\begin{aligned}\vartheta &: [0,R]^2 \to [0,\pi] \\ \vartheta(r,s) &= \begin{cases} \arccos\left(\frac{\cosh r \cosh s - \cosh R}{\sinh r \sinh s}\right), & \text{for } r,s \in (0,R], r+s \geq R \\ \pi, & \text{for } r,s \in [0,R], r+s < R \end{cases} \\ &= \text{ maximal angle between two adjacent vertices} \\ &\quad\;\; \text{with radial coordinates } r \text{ and } s.\end{aligned}$$

(Formally, we can also think of the KPKVB model as a probability distribution on the set of n-vertex graphs (with vertex set $[n] = \{1, \ldots, n\}$), where we associate with each vertex an auxiliary random vector (indicating the location of the vertex in the hyperbolic plane), s.t. the edge set is the 'deterministic geometric transformation' of the random vectors; note that two or more vertices can coincide in location, but this happens with probability zero).

In other words, given the parameters $n \in \mathbb{N}_0$, $0 < \nu < n$ and $\alpha > 0$, we derive the parameter $R = 2\log(n/\nu)$ and obtain the KPKVB model by sampling n points independently and uniformly in a disk with radius R around the origin of the hyperbolic plane with curvature $-\alpha^2$ and place an edge between a pair of vertices if their hyperbolic distance measured at a curvature of -1 is at most R (using the previously obtained polar coordinates of the vertices).

See Figure 1.1 for simulations of the KPKVB model with different values for ν and Figure 1.2 for simulations with different values for α. See Figure 1.3 for an illustration of the adjacency rule, resp. the neighbourhood ball of a vertex. Note that in these and ensuing simulation plots of the model, the distances differ from our usual (Euclidean) intuition. Roughly speaking, the distances are larger towards the boundary of the disk and behave a bit like the sum of the radial coordinates minus a correction term that depends on the angular distance between the two points [36].

The intuitive interpretation of the parameters is as follows: Clearly, n is the number of vertices. The parameter α determines how the points are distributed within the disk.

For $\alpha \to 0$ (and R fixed), $g_{\alpha,R}(r,\theta) \to \frac{r}{\pi R^2}$ which is the uniform distribution in the Euclidean disk with radius R (which has curvature zero); note however that in the KPKVB model, the distances are measured at a curvature of -1 for all $\alpha > 0$. For $\alpha \to \infty$ (and R fixed), the distribution given by $g_{\alpha,R}$ tends to the uniform distribution of the circle (curve) with radius R.

If $\alpha = 1$, the distribution of (r,θ) given by (1.1) is the uniform distribution on $\mathcal{D}_R$. For general $\alpha \in (0,\infty)$ Krioukov et al. [36] call the distribution (1.1) the *quasi-uniform* distribution on $\mathcal{D}_R$.

The parameter ν influences the radius R of the disk (which is also used as the adjacency threshold) in such a way that for $\alpha > \frac{1}{2}$, the average degree tends in probability to a finite positive constant [36].

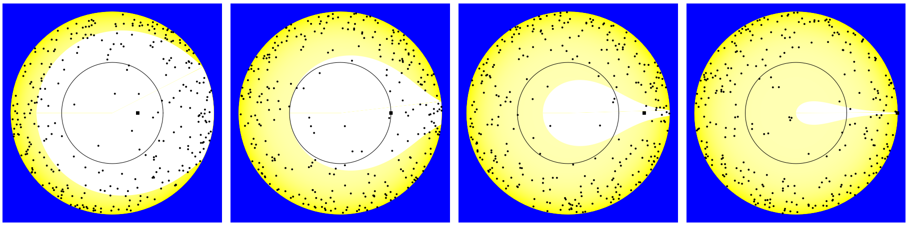

Figure 1.3: Simulations of the KPKVB model with $n = 300$, $\alpha = 0.6$, $\nu = 1$, indicating the neighbourhood ball (in white) for a vertex with radial coordinate $\frac{1}{4}R$, $\frac{1}{2}R$, $\frac{3}{4}R$, R (from left to right).

1.2.1 Previous results

Krioukov et al. [36] observed that the distribution of the degrees in $G(n; \alpha, \nu)$ follows a power-law with exponent $2\alpha + 1$, for $\alpha \in (1/2, \infty)$. For a certain range of degrees $k = k_n$, this was verified rigorously by Gugelmann et al. in [30]. Note that for $\alpha \in (1/2, 1)$, the exponent of the power-law is between 2 and 3, which is in line with numerous observations in networks which arise in applications (see for example [4]). In addition, Krioukov et al. observed, and Gugelmann et al. proved rigorously, that the (local) clustering coefficient of the graph stays bounded away from zero a.a.s. (we will give the definition of local clustering in Section 1.5). Here and in the rest of this thesis we use the following notation: If $(E_n)_{n\in\mathbb{N}}$ is a sequence of events then we say that E_n occurs *asymptotically almost surely (a.a.s.)*, if $\mathbb{P}(E_n) \to 1$ as $n \to \infty$.

Krioukov et al. [36] also observed that the average degree of $G(n; \alpha, \nu)$ is determined via the parameter ν for $\alpha \in (1/2, \infty)$. This was rigorously verified in Gugelmann et at. [30] too. In particular, Gugelmann et at. [30] proved that the average degree tends to $\frac{2\alpha^2\nu}{\pi(\alpha-\frac{1}{2})^2}$ in probability.

In Bode, Fountoulakis and Müller [11], it was established that $\alpha = 1$ is the critical point for the emergence of a giant component in $G(n; \alpha, \nu)$. In particular, if $\alpha \in (0, 1)$, the fraction of vertices contained in the largest component is bounded away from 0 a.a.s., whereas if $\alpha \in (1, \infty)$, the largest component is sublinear in n a.a.s. For $\alpha = 1$, the component structure depends on ν. If ν is large enough, then a giant component exists a.a.s., but if ν is small enough, then a.a.s. all components are sublinear.

In Fountoulakis and Müller [25], this picture is sharpened. There, it is shown that the fraction of vertices belonging to the largest component converges in probability to a constant which depends on α and ν. For $\alpha = 1$, the existence of a critical value $\nu_0 \in (0, \infty)$ is established such that when ν crosses ν_0 a giant component emerges a.a.s. [25]. In [32] and [33], Kiwi and Mitsche considered the size of the second largest component and showed that if $\alpha \in (\frac{1}{2}, 1)$, a.a.s., the second largest component has polylogarithmic order with exponent $\frac{1}{\alpha-\frac{1}{2}}$.

Apart from the degree distribution, clustering and component sizes, the graph distances in this model have also been considered. In [32] and [26], a.a.s. polylogarithmic upper and lower bounds on the diameter of the largest component are shown, and in [42], these were sharpened to show that $\log n$ is the correct order of the diameter. Furthermore, in [2] it is shown that for $\alpha \in (1/2, 1)$ the largest component is what is called an *ultra-small world*: it exhibits doubly logarithmic typical distances.

Results on the global clustering coefficient were obtained in [17], and on the evolution of graphs on more general spaces with negative curvature in [24]. The spectral gap of the Laplacian of this model was studied in [31].

In [12], Bode, Fountoulakis and Müller showed that $\alpha = 1/2$ is the critical value for connectivity: that is, if $\alpha \in (0, 1/2)$, then $G(n; \alpha, \nu)$ is a.a.s. connected, whereas $G(n; \alpha, \nu)$ is a.a.s. disconnected if $\alpha \in (1/2, \infty)$. The second half of this statement is in fact already immediate from the results of Gugelmann et al. [29]: there it is shown that for $\alpha > 1/2$, a.a.s., there are linearly many isolated vertices. For $\alpha = 1/2$, the probability of connectivity tends to a limiting value that is a function of ν which is continuous and non-decreasing and which equals one if and only if $\nu \geq \pi$.

1.3 Perfect matchings and Hamilton cycles

A *Hamilton cycle* in a graph is a closed path which contains all vertices of the graph. A graph is called *Hamiltonian* if it contains at least one Hamilton cycle. A *matching* is a set of edges that do not share endpoints and a *perfect matching* is a matching that covers all vertices of the graph.

Hamilton cycles and perfect matchings are classical topics in graph theory. The origin of the study of Hamilton cycles is usually traced to William Rowan Hamilton, although the topic had been studied before [10]. Hamilton labeled the vertices of a dodecahedron with different city names and asked for a round trip that visits each city exactly once. Finding a Hamilton cycle in a graph or solving the related traveling salesman problem are computationally hard. Historically the existence of Hamilton cycles and perfect matchings has been a central theme in the theory of random graphs as well. For instance, for the Erdős-Rényi model (also called the binomial random graph), the limit probability for having a Hamilton cycle has been derived [34,35,45]. In the random graph process (where in each step a new edge is added independently at random), a.a.s. the evolving graph turns Hamiltonian at the same time as it attains minimum degree at least two [3, 13]. In the context of random geometric graphs in the Euclidean plane, analogous results have been obtained [8, 21, 41]. The emergence of Hamilton cycles was also considered in other models, including the preferential attachment model [27] and the random d-regular graph model [46].

As one of the contributions of this thesis, we explore the existence of Hamilton cycles and perfect matchings in the KPKVB model $G(n; \alpha, \nu)$, the proofs can be found in Chapter 2. In light of the result on isolated vertices mentioned above, the question is non-trivial only for $\alpha \leq \frac{1}{2}$ (as for $\alpha > \frac{1}{2}$, the existence of isolated vertices implies that there is neither a perfect matching nor a Hamilton cycle). A perfect matching trivially cannot exist if n is odd. For this reason we find it convenient to switch to considering *near perfect matchings* from now on. That is, matchings that cover all but at most one vertex. (So if n is even a near perfect matching is the same as a perfect matching; and the existence of a Hamilton cycle implies the existence of a near perfect matching.)

We show that in the regime $\alpha < \frac{1}{2}$, a.a.s., the existence of a Hamilton cycle as well as of a (near) perfect matching has a non-trivial phase transition in ν:

Theorem 1.3.1. *For all positive real $\alpha < \frac{1}{2}$, there are constants $\nu_0 = \nu_0(\alpha)$ and $\nu_1 = \nu_1(\alpha)$ such that the following hold. For all $0 < \nu < \nu_0$, the random graph $G(n;\alpha,\nu)$ a.a.s. does not have a near perfect matching (and consequently no Hamilton cycle either). For all $\nu > \nu_1$, $G(n;\alpha,\nu)$ a.a.s. has a Hamilton cycle (and hence also a near perfect matching).*

To our knowledge, this is the first time this problem is considered for the $G(n;\alpha,\nu)$ model. Note that in the theorem above, we must have $\nu_0 \leq \nu_1$. We conjecture that the dependence on ν is sharp, i.e. $\nu_0 = \nu_1 =: \nu_c$.

Conjecture 1.3.2. *For every $0 < \alpha < \frac{1}{2}$, there exists a critical $\nu_c = \nu_c(\alpha) > 0$ such that if $\nu < \nu_c$, a.a.s. $G(n;\alpha,\nu)$ has no near perfect matching, whereas if $\nu > \nu_c$, then a.a.s. $G(n;\alpha,\nu)$ has a Hamilton cycle.*

A natural question to ask is what happens in the case $\alpha = \frac{1}{2}$. Does there exist ν large enough so that the graph a.a.s. becomes Hamiltonian in this case as well?

It would also be interesting to explore the relation of Hamiltonicity with the property of 2-connectivity. Recall that a graph is 2-connected if it has at least 3 vertices and it remains connected when removing a single vertex. Every Hamiltonian graph is 2-connected, but not vice versa in general. If the above conjecture is true, is there a similar behaviour for the property of 2-connectivity? If yes, are the corresponding critical constants ν_c equal?

1.4 Degree distribution

The degree of a vertex denotes the number of neighbours, i.e. the number of vertices that are adjacent to the given vertex. The degree distribution considers the relative frequency of a degree k. Previously, Gugelmann et al. [29] showed that the degree distribution follows a power-law with exponent $2\alpha + 1$ for all sequences k_n with $0 \leq k_n \leq n^{\delta'}$, where $\delta' < \min\left(\frac{2\alpha-1}{4(2\alpha+1)\alpha}, \frac{2(2\alpha-1)}{5(2\alpha+1)}\right) < \frac{1}{2\alpha+1}$. As a contribution of this thesis, we extend this result: we show that the power-law holds all the way up to $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$, that there are no vertices of degree exactly k_n for larger scaling $k_n \gg n^{\frac{1}{2\alpha+1}}$ a.a.s. and that there is a Poisson limiting

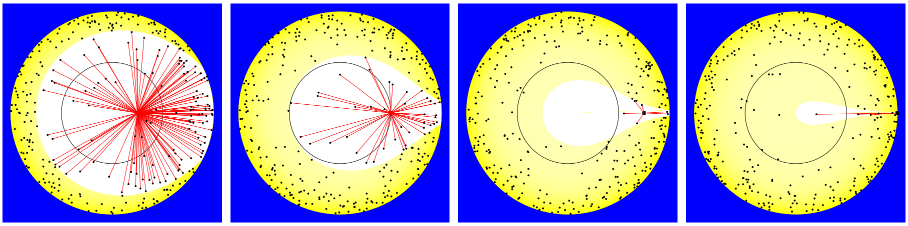

Figure 1.4: As we move a vertex away from the boundary towards the centre (from right to left), its expected degree increases exponentially, but the probability mass of having a point at that distance from the boundary also decreases exponentially. This results in the overall power-law of the degree distribution of the KPKVB model.

distribution in the boundary case. The proofs can be found in Chapter 3. See Figure 1.4 for an illustration of our results regarding the degree distribution of the KPKVB model. Let $Po(\lambda)$ denote a Poisson random variable with expectation $\lambda > 0$. Let

$$\Gamma^{+}(a,b) := \int_{b}^{\infty} t^{a-1}e^{-t}\mathrm{d}t$$

denote the upper incomplete gamma function.

Theorem 1.4.1. *Let* $\alpha > \frac{1}{2}$. *Let* $\xi = \frac{4\alpha\nu}{\pi(2\alpha-1)}$.

Let $N_n(k)$ *denote the number of degree* k *vertices in the KPKVB model* $G(n;\alpha,\nu)$ *and consider a sequence of integers* $(k_n)_n$ *with* $0 \leq k_n \leq n-1$.

1. *If* $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$ *as* $n \to \infty$, *then a.a.s.*

$$N_n(k_n) = (1+o(1))np_{k_n},$$

where

$$p_{k_n} = \frac{2\alpha\xi^{2\alpha}\Gamma^{+}(k_n - 2\alpha, \xi)}{k_n!}.$$

2. *If* $k_n = (1+o(1))cn^{\frac{1}{2\alpha+1}}$ *for some fixed* $c > 0$, *then*

$$N_n(k_n) \xrightarrow[n\to\infty]{d} \mathrm{Po}(2\alpha\xi^{2\alpha}c^{-(2\alpha+1)}).$$

3. *If* $k_n \gg n^{\frac{1}{2\alpha+1}}$, *then a.a.s.* $N_n(k_n) = 0$.

Note that in the theorem above, $p_{k_n} = (1+o(1))2\alpha\xi^{2\alpha}k_n^{-(2\alpha+1)}$ as $k_n \to \infty$.

1.5 Clustering

In the literature there are two conceptually distinct definitions of the *clustering coefficient*. One of these, sometimes called the *global* clustering coefficient, measures the extent to which the adjacency relation is transitive (i.e. if a is adjacent to b and b is adjacent to c, what is the probability that a is adjacent to c?). It is defined as three times the ratio of the number of triangles to the number of paths of length two in the graph. Results for this version of the clustering coefficient in the KPKVB model were obtained by Candellero and Fountoulakis [17] and for the evolution of graphs on more general spaces with negative curvature by Fountoulakis in [24].

We will study the other notion of clustering, the one which is also considered by Krioukov et al. [36] and Gugelmann et al. [29]. It is sometimes called the *local* clustering coefficient, although we should point out that Gugelmann et al. actually call it the global clustering coefficient in their paper. It is a number between zero and one measuring the extent to which the neighbourhood of a vertex resembles a clique. More precisely, for a graph G and a vertex $v \in V(G)$ we define the clustering coefficient *of* v as

$$c(v) := \begin{cases} \dfrac{1}{\binom{\deg(v)}{2}} \displaystyle\sum_{u,w \sim v} \mathbb{1}_{\{uw \in E(G)\}}, & \text{if } \deg(v) \geq 2, \\ \\ 0, & \text{otherwise,} \end{cases}$$

where $E(G)$ denotes the edge set of G and $\deg(v)$ is the degree of vertex v. That is, provided v has degree at least two, $c(v)$ equals the number of edges that are actually present between the neighbours of v divided by the number of edges that could possibly be present between the neighbours given the degree of v. The clustering coefficient of G is now defined as the average of $c(v)$ over all vertices v:

$$c(G) := \frac{1}{|V(G)|} \sum_{v \in V(G)} c(v).$$

As mentioned above, Gugelmann et al. [29] have established that $c(G(n; \alpha, \nu))$ is non-vanishing a.a.s., but they left open the question of convergence. As a contribution of this thesis, in Theorem 1.5.1 below, we address this question and establish that the clustering coefficient indeed converges in probability to a constant γ that we give explicitly as a closed-form expression involving α, ν and several classical special functions.

In addition to the clustering coefficient, we shall also be interested in the *clustering function* which assigns to every non-negative integer k the average of the local clustering coefficients over all vertices of degree k, i.e.

$$c(k) = c(k; G) := \begin{cases} \dfrac{1}{N(k)} \displaystyle\sum_{\substack{v \in V(G), \\ \deg(v)=k}} c(v), & \text{if } N(k) \geq 1, \\ 0, & \text{otherwise}, \end{cases} \tag{1.2}$$

where $N(k)$ denotes the number of vertices of degree exactly k in G. We remark that, while it might seem natural to consider $c(k)$ to be "undefined" when $N(k) = 0$, we prefer to use the above definition for technical convenience. This way $c(k; G(n; \alpha, \nu))$ is a standard random variable and we can for instance compute its moments without any issues.

A general expression of the clustering function for KPKVB random graphs is given in Krioukov et al. [36, Equation (59)]. The authors state that as k tends to infinity, the clustering function decays as k^{-1}. They based this claim on observations (Figure 8 in [36]) in experiments on the infrastructure of the Internet obtained in [19]. Despite these interesting observations and the attention the KPKVB model has generated since then, the behaviour of the clustering function in KPKVB random graphs had not been completely determined. In particular, it had not been established whether it converges as $n \to \infty$ to some suitable limit function, nor how $c(k; G(n; \alpha, \nu))$ scales with k. As a contribution of this thesis, Theorems 1.5.2, 1.5.3 and Proposition 1.5.4 below settle these questions. Theorem 1.5.2 shows that for each fixed k, the value $c(k; G(n; \alpha, \nu))$ converges in probability to a constant $\gamma(k)$ that we again give explicitly as a closed-form expression involving α, ν and several classical special functions. Theorem 1.5.3 extends this result to increasing sequences satisfying $k \ll n^{1/(2\alpha+1)}$. Proposition 1.5.4 clarifies the asymptotic behavior of the limiting function $\gamma(k)$, as $k \to \infty$. This depends on the parameter α, and $\gamma(k)$ only scales with k^{-1} if $\alpha > 3/4$, which corresponds to the exponent of the degree distribution exceeding $5/2$. So in particular, our findings disprove the abovementioned claim of Krioukov et al. [36]. All the proofs can be found in Chapter 4.

Notation

Throughout the rest of the thesis, we will use the following notations. We set

$$\xi := \frac{4\alpha\nu}{\pi(2\alpha - 1)}.$$

We write

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t$$

for the gamma function,

$$\Gamma^+(a, b) := \int_b^\infty t^{a-1} e^{-t} \mathrm{d}t$$

for the upper incomplete gamma function,

$$B(a,b) := \int_0^1 u^{a-1}(1-u)^{b-1}\mathrm{d}u = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

for the beta function and

$$B^-(x;a,b) := \int_0^x u^{a-1}(1-u)^{b-1}\mathrm{d}u$$

for the lower incomplete beta function. We write $U(a,b,z)$ for the hypergeometric U-function (also called Tricomi's confluent hypergeometric function), which has the integral representation

$$U(a,b,z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zt}t^{a-1}(1+t)^{b-a-1}dt,$$

see [23, p.255 Equation (2)], and let $G^{m,\ell}_{p,q}\left(z\middle|\begin{matrix}\mathbf{a}\\ \mathbf{b}\end{matrix}\right)$ denote Meijer's G-Function [39], see Appendix A for more details.

For a sequence $(X_n)_n$ of random variables, we write $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ to denote that X_n converges in probability to X, as $n \to \infty$.

1.5.1 The clustering coefficient

Theorem 1.5.1. *Let $\alpha > \frac{1}{2}$, $\nu > 0$ be fixed. Writing $G_n := G(n;\alpha,\nu)$, we have*

$$c(G_n) \xrightarrow[n\to\infty]{\mathbb{P}} \gamma,$$

where γ is defined for $\alpha \neq 1$ as

$$\begin{aligned}\gamma = {} & \frac{2+4\alpha+13\alpha^2-34\alpha^3-12\alpha^4+24\alpha^5}{16(\alpha-1)^2\alpha(\alpha+1)(2\alpha+1)} + \frac{2^{-1-4\alpha}}{(\alpha-1)^2}\\
& + \frac{(\alpha-1/2)(B(2\alpha,2\alpha+1)+B^-(1/2;1+2\alpha,-2+2\alpha))}{2(\alpha-1)(3\alpha-1)}\\
& + \frac{\xi^{2\alpha}\left(\Gamma^+(1-2\alpha,\xi)+\Gamma^+(-2\alpha,\xi)\right)}{4(\alpha-1)}\\
& + \frac{\xi^{2\alpha+2}\alpha(\alpha-1/2)^2\left(\Gamma^+(-2\alpha-1,\xi)+\Gamma^+(-2\alpha-2,\xi)\right)}{2(\alpha-1)^2}\\
& - \frac{\xi^{2\alpha+1}\alpha(2\alpha-1)\left(\Gamma^+(-2\alpha,\xi)+\Gamma^+(-2\alpha-1,\xi)\right)}{(\alpha-1)}\\
& - \frac{\xi^{6\alpha-2}2^{-4\alpha}(3\alpha-1)\left(\Gamma^+(-6\alpha+3,\xi)+\Gamma^+(-6\alpha+2,\xi)\right)}{(\alpha-1)^2}\end{aligned}$$

$$
\begin{aligned}
&- \frac{\xi^{6\alpha-2}(\alpha-1/2)B^{-}(1/2;1+2\alpha,-2+2\alpha)\left(\Gamma^{+}(-6\alpha+3,\xi)+\Gamma^{+}(-6\alpha+2,\xi)\right)}{(\alpha-1)} \\
&- \frac{e^{-\xi}\Gamma(2\alpha+1)\left(U(2\alpha+1,1-2\alpha,\xi)+U(2\alpha+1,2-2\alpha,\xi)\right)}{4(\alpha-1)} \\
&+ \frac{\xi^{6\alpha-2}\Gamma(2\alpha+1)\left(G^{3,0}_{2,3}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\3-4\alpha,-6\alpha+2,0\end{matrix}\right)+G^{3,0}_{2,3}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\3-4\alpha,-6\alpha+3,0\end{matrix}\right)\right)}{4(\alpha-1)},
\end{aligned}
$$

and for $\alpha = 1$ as the $\alpha \to 1$ limit of the above expression.

A plot of γ can be found in Figure 1.5.

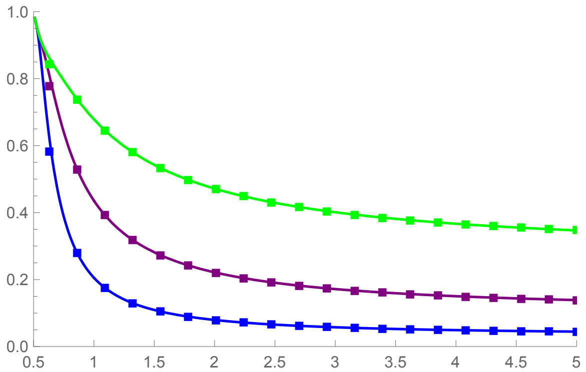

Figure 1.5: Plot of γ for α varying from 0.5 to 5 on the horizontal axis and for $\nu = \frac{1}{2}$ (blue), $\nu = 1$ (purple), $\nu = 2$ (green); simulations (squares in corresponding colour) with $n = 10000$ and 100 repetitions.

In the above expression for γ, a factor $\alpha - 1$ occurs in the denominator of each term, but we will see that this corresponds to a removable singularity. We have not been able to find a closed-form expression in terms of known functions in the case $\alpha = 1$, but in Section 4.1.2 we do provide an explicit expression involving integrals.

1.5.2 The clustering function

Figure 1.6 shows how the clustering changes with the degree and how vertices of a given degree are concentrated around a particular height. Figure 1.7 illustrates how the clustering coefficient of a (fixed additional) vertex changes as we alter its radial coordinate.

Theorem 1.5.2. *Let $\alpha > \frac{1}{2}$, $\nu > 0$ and $k \geq 2$ be fixed. Writing $G_n := G(n;\alpha,\nu)$, we have*

$$c(k;G_n) \xrightarrow[n\to\infty]{\mathbb{P}} \gamma(k),$$

where $\gamma(k)$ is defined for $\alpha \neq 1$ as

$$\begin{aligned}&\frac{1}{8\alpha(\alpha-1)\Gamma^+(k-2\alpha,\xi)}\left(-\Gamma^+(k-2\alpha,\xi) - 2\frac{\alpha(\alpha-1/2)^2\xi^2\Gamma^+(k-2\alpha-2,\xi)}{(\alpha-1)}\right.\\ &+8\alpha(\alpha-1/2)\xi\Gamma^+(k-2\alpha-1,\xi)\\ &+4\xi^{4\alpha-2}\Gamma^+(k-6\alpha+2,\xi)\left(\frac{2^{-4\alpha}(3\alpha-1)}{(\alpha-1)} + (\alpha-1/2)B^-(1/2;1+2\alpha,-2+2\alpha)\right)\\ &+\xi^{k-2\alpha}\Gamma(2\alpha+1)e^{-\xi}U(2\alpha+1,1+k-2\alpha,\xi)\\ &\left.-\xi^{4\alpha-2}\Gamma(2\alpha+1)G^{3,0}_{2,3}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\ 3-4\alpha,-6\alpha+k+2,0\end{matrix}\right)\right)\end{aligned}$$

and for $\alpha = 1$ as the $\alpha \to 1$ limit of the above expression.

A plot of $\gamma(k)$ can be found in Figure 1.8. Again, we remark that the above expression for $\gamma(k)$ appears to have a singularity at $\alpha = 1$, but this will turn out to be a removable singularity. Again, we have not been able to find a closed-form expression in terms of known functions in the case when $\alpha = 1$, but in Section 4.1.2 we do provide an explicit expression involving integrals.

Theorem 1.5.2 in fact generalises to increasing sequences $(k_n)_{n\geq 1}$.

Theorem 1.5.3. *Let $\alpha > \frac{1}{2}, \nu > 0$ be fixed and let $(k_n)_n$ be a sequence of non-negative integers satisfying $1 \ll k_n \ll n^{\frac{1}{2\alpha+1}}$. Then, writing $G_n := G(n;\alpha,\nu)$, we have*

$$\frac{c(k_n;G_n)}{\gamma(k_n)} \xrightarrow{\mathbb{P}} 1,$$

as $n \to \infty$, where $\gamma(\cdot)$ is as in Theorem 1.5.2. Note that this might alternatively be written as $c(k_n;G_n) = (1+o(1))\gamma(k_n)$ a.a.s., using notation that is common in the random graphs community.

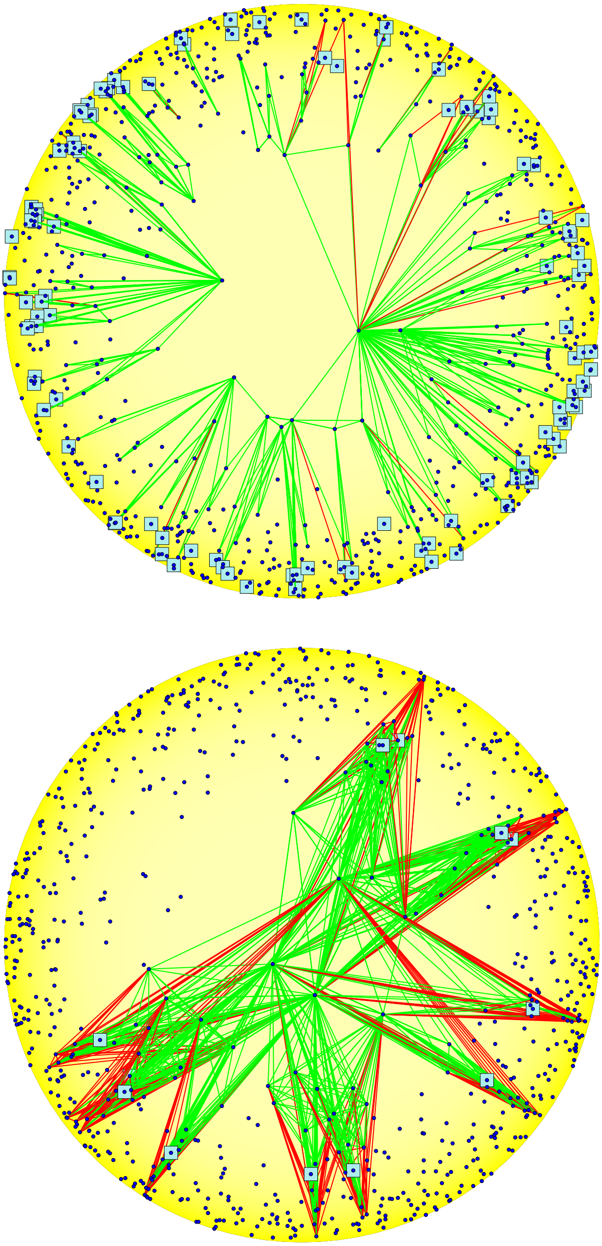

Figure 1.6: This figure shows simulations of the KPKVB model with $n = 800$ vertices, $\alpha = 0.6$ and $\nu = 1$, with all vertices of degree 4 resp. 16 highlighted. It illustrates that vertices of a given degree concentrate at a particular height (distance from the boundary of the disk) and that the clustering function is decreasing in the degree k.

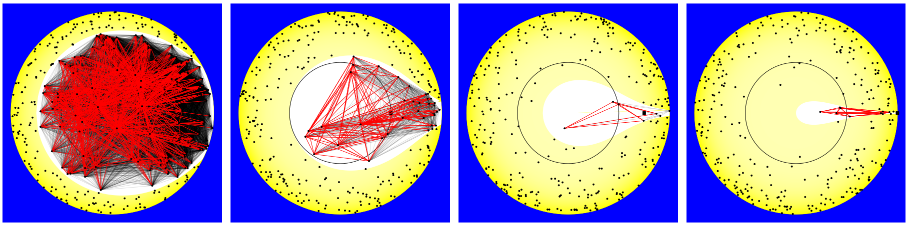

Figure 1.7: Simulations of the KPKVB model with $n = 300$ vertices, $\alpha = 0.6$, $\nu = 1$, illustrating how the clustering function changes as we move a vertex from the boundary towards the centre of the disk (from right to left).

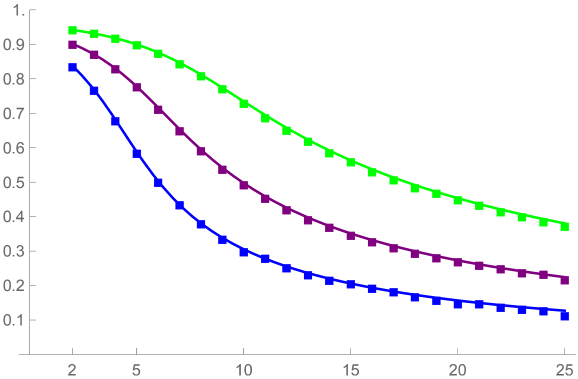

Figure 1.8: Plot $\gamma(k)$ for k varying from 2 to 25 on the horizontal axis, for $\alpha = 0.8$ and $\nu = \frac{1}{2}$ (blue), $\nu = 1$ (purple), $\nu = 2$ (green); simulations (squares in corresponding colour) with $n = 10000$ and 100 repetitions.

Scaling of $\gamma(k)$

To clarify the scaling behaviour of $\gamma(k)$ with k we offer the following result.

Proposition 1.5.4. *As $k \to \infty$, we have*

$$\gamma(k) = (1 + o(1)) \cdot \begin{cases} \frac{8\alpha\nu}{\pi(4\alpha-3)} \cdot k^{-1}, & \text{if } \alpha > \frac{3}{4}, \\ \frac{6\nu}{\pi} \cdot \frac{\log(k)}{k}, & \text{if } \alpha = \frac{3}{4}, \\ c_\alpha \cdot k^{2-4\alpha}, & \text{if } \frac{1}{2} < \alpha < \frac{3}{4}, \end{cases}$$

where $c_{\alpha,\nu} := \left(\frac{3\alpha-1}{2^{4\alpha+1}\alpha(\alpha-1)^2} + \frac{(\alpha-\frac{1}{2})B^{-}(\frac{1}{2};2\alpha+1,2\alpha-2)}{2(\alpha-1)\alpha} - \frac{B(2\alpha,3\alpha-4)}{4(\alpha-1)} \right) \cdot \xi^{4\alpha-2}$.

Note that Theorem 1.5.3 implies that as k grows, the clustering function of the KPKVB model scales with $\gamma(k)$, whose scaling is given in the above result. In particular, this contradicts the scaling claimed in Krioukov et al. [36] for $\alpha \leq \frac{3}{4}$, and confirms it only for $\alpha > \frac{3}{4}$.

We remark that simultaneously and independently Stegehuis, van der Hofstad and van Leeuwaarden [47] used a completely different technique to obtain a similar result on the $k \to \infty$ scaling of the clustering function in the KPKVB model. Their statement has a case distinction depending on the exponent of the power-law of the degree distribution. Proposition 1.5.4 confirms their result and additionally yields the leading constants of the scaling. While the focus of [47] lies on the explanation of the big picture and derivations with applicability to many random graph models, we focus here on the careful treatment of technical details and the rigor and clarity of the mathematical exposition.

In [47], they introduce a variational principle, i.e. they ask: among all triples consisting of a vertex with degree k, a vertex with degree d_1 and a vertex with degree d_2, what are the values of d_1 and d_2 that maximize the probability that the three vertices form a triangle or, in other words, are pairwise adjacent? Our method, on the other hand, consists of showing convergence to an infinite limit model, where we compute the clustering coefficient of a typical vertex via some integral. A crucial tool for deriving the limiting integral is the Campbell-Mecke formula. We will give a more detailed introduction and explanation of the infinite limit model, the Campbell-Mecke formula and other useful ideas in Section 1.6. A very closely related way of proving the convergence of the clustering function would be via local weak convergence. This concept was introduced by Benjamini and Schramm [9] and independently by Aldous and Steele [5] to formalize the idea that in a sequence of graphs G_n, as $n \to \infty$, the local structure around a typical vertex in G_n becomes more and more similar to the local structure around the root of a (typically infinite) graph G_∞. Our arguments follow these ideas without mentioning local weak convergence explicitly. For an introduction of local weak convergence in the context of random graphs, we refer to [50].

1.5.3 Observations

There are a few other observations to be made regarding our results.

Near maximum scaling for k_n. Our results for the clustering function in the KPKVB model are valid for any sequence of non-negative integers $k_n \to \infty$ such that $k_n \ll n^{\frac{1}{2\alpha+1}}$. Although one would like to have results for any sequence $k_n \leq n-1$, it turns out that $n^{\frac{1}{2\alpha+1}}$ is the near maximum scaling for which the clustering function is non-zero: by Theorem 1.4.1, if $k_n \gg n^{\frac{1}{2\alpha+1}}$, then a.a.s., $N_n(k_n) = 0$ and hence also $c(k_n; G_n) = 0$.

Transition in scaling at $\alpha = 3/4$. It follows from Proposition 1.5.4 that there is a transition in the scaling of the local clustering function at $\alpha = 3/4$. This corresponds to an exponent $5/2$ for the power-law of the probability mass function of the degree distribution. Interestingly, a similar transition point has also been observed during the analysis of a bidirectional shortest path algorithm on the KP-

KVB model and for both majority vote models [18] and flocking dynamics [40] on networks with power-law degree distributions. In Stegehuis, van der Hofstad and van Leeuwaarden [47], it is observed that the main contribution to the clustering function changes at $\alpha = \frac{3}{4}$: for $\alpha < \frac{3}{4}$, i.e. for a power-law exponent $< \frac{5}{2}$, the main contribution to the clustering function stems from neighbours with degree proportional to k, whereas for $\alpha > \frac{3}{4}$, i.e. for a power-law exponent $> \frac{5}{2}$, it stems from neighbours with constant degree. We think that further investigations on the transition point $\alpha = \frac{3}{4}$ could still be a fertile ground for new interesting insights.

1.6 Related models and tools

There are several ways in which the KPKVB model can be approximated and which give rise to useful mathematical properties. First of all, instead of using a fixed number n of vertices, we can draw a Poisson distributed random variable N and then sample that many vertices for the graph. This has the effect that the vertex set turns into a Poisson point process, which implies that the number of vertices in disjoint subsets of the disk will be independent Poisson random variables. In the context of the Poisson point process, we can use the Campbell-Mecke formula: if we want to compute the expectation of some quantity when summed over all vertices of the graph (where this quantity may still depend on the entire point process resp. graph), we can instead drop an additional vertex and compute the expectation of the quantity for this additional vertex.

Before we turn to other approximations, we introduce new coordinates, i.e. instead of polar coordinates $u = (r, \theta)$ in the hyperbolic disk $\mathcal{D}$, we use Cartesian coordinates $p = (x, y)$ in the box $\mathcal{R} = (-\frac{\pi}{2}e^{R/2}, \frac{\pi}{2}e^{R/2}] \times (0, R] \subset \mathbb{R}^2$. In other words, we apply the coordinate transformation (described below) to the vertices given in polar coordinates and we compute the adjacency for vertices given in the Cartesian coordinates by applying the inverse coordinate transformation and then using the original adjacency rule.

Using the new coordinates, if $\alpha > \frac{1}{2}$, we will make two further approximations: firstly to the probability density resp. intensity measure used to sample the vertices and secondly to the formula for determining the adjacency of two vertices. Finally, if we let n tend to infinity, these approximations lead to a limit model which is very convenient for practical computations (roughly speaking, we are replacing the hyperbolic functions by exponential functions).

After giving more details on the related models (which we will do more conveniently in the reverse order of how we just introduced and motivated them), we will state the versions of the Campbell-Mecke formula and of the Chernoff bound that we will be using repeatedly throughout the thesis.

1.6.1 The infinite limit model G_∞

We start by recalling the definition of the infinite limit model from Fountoulakis and Müller [25]. Let $\mathcal{P} = \mathcal{P}_{\alpha,\nu}$ be a Poisson point process on $\mathbb{R}^2$ with intensity function $f = f_{\alpha,\nu}$ given by

$$f(x, y) = \frac{\alpha\nu}{\pi} e^{-\alpha y} \cdot \mathbb{1}_{\{y>0\}}. \tag{1.3}$$

The *infinite limit model* $G_\infty = G_\infty(\alpha, \nu)$ has vertex set $\mathcal{P}$ and edge set such that

$$pp' \in E(G_\infty) \iff |x - x'| \le e^{\frac{y+y'}{2}},$$

for $p = (x, y), p' = (x', y') \in \mathcal{P}$.

For any point $p \in \mathbb{R} \times (0, \infty)$, we write $\mathcal{B}_\infty(p)$ to denote the *ball* around p, i.e.

$$\mathcal{B}_\infty(p) = \{p' \in \mathbb{R} \times (0, \infty) : |x - x'| \le e^{\frac{y+y'}{2}}\}. \tag{1.4}$$

With this notation we then have that $\mathcal{B}_\infty(p) \cap \mathcal{P}$ denotes the set of neighbours of a vertex $p \in G_\infty$. We will denote the intensity measure of the Poisson process $\mathcal{P}$ by $\mu = \mu_{\alpha,\nu}$, i.e. for every Borel-measurable subset $S \subseteq \mathbb{R} \times (0, \infty)$ we have $\mu(S) = \int_S f(x, y)\,\mathrm{d}x\,\mathrm{d}y$.

1.6.2 The finite box model G_{box}

For the definition of the finite graph, recall that in the definition of the KPKVB model we set $R = 2\log(n/\nu)$. We consider the box $\mathcal{R} = (-\frac{\pi}{2}e^{R/2}, \frac{\pi}{2}e^{R/2}] \times (0, R]$ in $\mathbb{R}^2$. Then the *finite box model* $G_{\text{box}} := G_{\text{box}}(n; \alpha, \nu)$ has vertex set $\mathcal{V}_{\text{box}} := \mathcal{P} \cap \mathcal{R}$ and edge set such that

$$pp' \in E(G_{\text{box}}(n; \alpha, \nu)) \iff |x - x'|_{\pi e^{R/2}} \le e^{\frac{y+y'}{2}},$$

where $|x|_r = \min(|x|, r - |x|)$ for $-r \le x \le r$. Using $|.|_{\pi e^{R/2}}$ instead of $|.|$ results in the left and right boundaries of the box $\mathcal{R}$ getting identified, which in particular makes the model invariant under horizontal shifts and reflections in vertical lines. The graph G_{box} can thus be seen as a subgraph of G_∞ induced on $\mathcal{V}_{\text{box}}$, with some additional edges caused by the identification of the boundaries.

Similar to the infinite graph, for a point $p \in \mathcal{R}$ we define the ball $\mathcal{B}_{\text{box}}(p)$ as

$$\mathcal{B}_{\text{box}}(p) = \left\{p' \in \mathcal{R} : |x - x'|_{\pi e^{R/2}} \le e^{\frac{y+y'}{2}}\right\}. \tag{1.5}$$

1.6.3 The Poissonized KPKVB model $G_{\mathbf{Po}}$

Imagine that we have an infinite supply of i.i.d. points $u_1, u_2, \ldots$ in the hyperbolic plane $\mathbb{H}$ chosen according to the (α, R)-quasi uniform distribution. In the standard KPKVB random graph $G(n; \alpha, \nu)$ we take $u_1, \ldots, u_n$ as our vertex set and add edges between points at hyperbolic distance at most $R = 2\log(n/\nu)$. In the *Poissonized* KPKVB random graph $G_{\text{Po}} := G_{\text{Po}}(n; \alpha, \nu)$, we instead take $N \stackrel{\mathrm{d}}{=} \text{Po}(n)$, a Poisson random variable with expectation n, independent of our i.i.d. sequence of points and let the vertex set be $u_1, \ldots, u_N$ and add edges according to the same rule as before. Equivalently, we could say that the vertex set consists of the points of a Poisson point process with intensity function ng, where g denotes the probability density of the (α, R)-quasi uniform distribution, i.e.

$$g(r, \theta) = \frac{\alpha \sinh(\alpha r)}{2\pi(\cosh(\alpha R) - 1)} \cdot \mathbb{1}_{\{0 \le r \le R, -\pi < \theta \le \pi\}}. \tag{1.6}$$

Working with the Poissonized model has the advantage that when we take two disjoint regions A, B then the number of points in A and the number of points in B are independent Poisson-distributed random variables. As we will see, and as is to be expected, switching to the Poissonized model does not significantly alter the limiting behaviour of the clustering coefficient and function or of the other properties we study.

1.6.4 Coupling $G_{\mathbf{Po}}$ and G_{box}

The following lemmas from Fountoulakis and Müller [25] establish a useful coupling between the Poissonized KPKVB random graph and the finite box model and relate the edge sets of the two graphs.

Lemma 1.6.1 (Fountoulakis-Müller [25, Lemma 27])**.** *Let $\mathcal{V}_{\text{Po}}$ denote the vertex set of $G_{Po}(n; \alpha, \nu)$ and $\mathcal{V}_{box}$ the vertex set of $G_{\text{box}}(n; \alpha, \nu)$. Define the map $\Psi : [0, R] \times (-\pi, \pi] \to \mathcal{R}$ by*

$$\Psi(r, \theta) = \left(\theta \frac{e^{R/2}}{2}, R - r\right). \tag{1.7}$$

Then there exists a coupling such that, a.a.s., $\mathcal{V}_{box} = \Psi(\mathcal{V}_{\text{Po}})$.

Roughly speaking, the fact that $\mathcal{V}_{\text{box}} = \Psi(\mathcal{V}_{\text{Po}})$ a.a.s. means that, when dealing with statements in probability, we can assume that G_{Po} and G_{box} have the same vertex sets, or in more detail that the locations of all vertices of G_{Po} and G_{box} fully agree, but are expressed in different coordinates.

In the remainder of this thesis, we will write $\mathcal{B}(p)$ to denote the image under Ψ of the ball of hyperbolic radius R around the point $\Psi^{-1}(p)$ for $p \in \mathcal{R}$, i.e.

$$\mathcal{B}(p) := \Psi\left(\left\{u \in \mathcal{D}_R : d_{\mathbb{H}}(\Psi^{-1}(p), u) \le R\right\}\right) \subset \mathcal{R}.$$

Under the map Ψ, a point $p = (x, y) \in \mathcal{R}$ corresponds to $u := \Psi^{-1}(p) = (2e^{-R/2}x, R - y)$.

By the hyperbolic rule of cosines, for two points $p = (x, y) = \Psi((r, \theta)), p' = (x', y') = \Psi((r', \theta')) \in \mathcal{R}$ we have that $p' \in \mathcal{B}(p)$ iff. either $r + r' \le R$ or $r + r' > R$ and

$$\cosh r \cosh r' - \sinh r \sinh r' \cos(|\theta - \theta'|_{2\pi}) \le \cosh(R),$$

This can be rephrased as $p' \in \mathcal{B}(p)$ iff. either $y + y' \ge R$ or $y + y' < R$ and

$$|x-x'|_{\pi e^{R/2}} \le \Phi(y,y') := \frac{1}{2} e^{R/2} \arccos\left(\frac{\cosh(R-y)\cosh(R-y') - \cosh R}{\sinh(R-y)\sinh(R-y')} \right). \tag{1.8}$$

The following lemma provides useful bounds on the function $\Phi(y,y')$. Note that in [25] the function Φ is written in terms of $r := R-y, r' := R-y'$.

Lemma 1.6.2 (Fountoulakis-Müller [25, Lemma 28]). *There exists a constant $K > 0$ such that, for every $\varepsilon > 0$ and for R sufficiently large, the following holds. For every $r, r' \in [\varepsilon R, R]$ with $y + y' < R$ we have that*

$$e^{\frac{1}{2}(y+y')} - Ke^{\frac{3}{2}(y+y')-R} \le \Phi(y,y') \le e^{\frac{1}{2}(y+y')} + Ke^{\frac{3}{2}(y+y')-R}. \tag{1.9}$$

Moreover:

$$\Phi(y,y') \ge e^{\frac{1}{2}(y+y')} \quad \text{if} \quad y, y' > K. \tag{1.10}$$

In its original formulation with radial coordinates (which we will also use), the same lemma reads as follows.

Lemma 1.6.3 (Fountoulakis-Müller [25], Lemma 28). *There exists a constant $K > 0$ such that for every $\epsilon > 0$ and R sufficiently large, the following holds: for every $r_1, r_2 \in [\epsilon R, R]$ with $r_1 + r_2 > R$, we have*

$$2e^{\frac{1}{2}(R-r_1-r_2)} - Ke^{\frac{3}{2}(R-r_1-r_2)} \le \vartheta(r_1,r_2) \le 2e^{\frac{1}{2}(R-r_1-r_2)} + Ke^{\frac{3}{2}(R-r_1-r_2)}.$$

Moreover, if $r_1, r_2 < R - K$, then $\vartheta(r_1,r_2) \ge 2e^{\frac{R-r_1-r_2}{2}}$.

A key consequence of Lemma 1.6.2 is that the coupling from Lemma 1.6.1 preserves edges between points whose heights are not too large.

Lemma 1.6.4 (Fountoulakis-Müller [25, Lemma 30]). *On the coupling space of Lemma 1.6.1 the following holds a.a.s.:*

1. *for any two points $p, p' \in \mathcal{V}_{box}$ with $y, y' \le R/2$,*
$$pp' \in E(G_{\text{box}}) \Rightarrow \Psi^{-1}(p)\Psi^{-1}(p') \in E(G_{Po}),$$
2. *for any two points $p, p' \in \mathcal{V}_{box}$ with $y, y' \le R/4$,*
$$pp' \in E(G_{\text{box}}) \iff \Psi^{-1}(p)\Psi^{-1}(p') \in E(G_{Po}).$$

Remark 1.6.5 (Notational convention for points). We will often be working with the finite box graph G_{box} or the infinite graph G_∞, whose vertices are points in $\mathbb{R}\times\mathbb{R}_+$. For any point $p \in \mathbb{R}\times\mathbb{R}_+$ we will always use $p = (x, y)$. When considering different points $p, p' \in \mathbb{R}\times\mathbb{R}_+$, we will use primed coordinates to refer to p', i.e. $p' = (x', y')$, and similar with subscripts, i.e. $p_i = (x_i, y_i)$.

1.6.5 The Campbell-Mecke formula

A very useful tool for analysing subgraph counts, and their generalizations, in the setting of the Poissonized random geometric graphs, and in particular the Poissonized KPKVB model and the box model is the *Campbell-Mecke formula.* We use a specific incarnation, which follows from the Palm theory of Poisson point processes on metric spaces, see [37]. For this, consider a Poisson point process $\mathcal{P}$ on some metric space S with intensity measure μ and let $\mathcal{N}$ denote the set of all possible point configurations in S, equipped with the σ-algebra of the process $\mathcal{P}$. Then, for any natural number k and measurable function $h : S^k \times \mathcal{N} \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\left[\sum_{p_1,\dots,p_k\in\mathcal{P}}^{\neq} h(p_1,\dots,p_k,\mathcal{P})\right] \qquad (1.11)\\ = \int_S \cdots \int_S \mathbb{E}\left[h(x_1,\dots,x_k,\mathcal{P}\cup\{x_1,\dots,x_k\})\right] \mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_k),\end{aligned}$$

where the sum is over all distinct points $p_1,\dots,p_k \in \mathcal{P}$.

1.6.6 Chernoff bounds

The Chernoff bounds are a collection of inequalities that are very useful and commonly studied in probability theory. Here, we will make use of the following version for Poisson random variables (see [44, Lemma 1.2]):

Lemma 1.6.6. *Let* $\mathrm{Po}(\mu)$ *denote a Poisson random variable with expectation* μ *and let* $H(x) = x\log(x) - x + 1$*. Then*

$$\mathbb{P}\left(\mathrm{Po}(\mu) \geq k\right) \leq e^{-\mu H\left(\frac{k}{\mu}\right)} \quad \textit{for all } k \geq \mu,$$
$$\mathbb{P}\left(\mathrm{Po}(\mu) \leq k\right) \leq e^{-\mu H\left(\frac{k}{\mu}\right)} \quad \textit{for all } k \leq \mu.$$

Note that in particular if $\frac{k}{\mu} \leq \frac{1}{2}$, then $H(\frac{k}{\mu}) \geq \frac{1}{2}(1-\ln 2) > 0$ (using that H is monotone decreasing in $(0,1)$). It follows from the above lemma that

$$\mathbb{P}\left(|\mathrm{Po}(\mu) - \mu| \geq x\right) \leq 2e^{-\frac{x^2}{2(\mu+x)}}. \qquad (1.12)$$

In particular, if $\mu_n \to \infty$, then, for any $C > 0$,

$$\mathbb{P}\left(|\mathrm{Po}(\mu_n) - \mu_n| \geq C\sqrt{\mu_n\log(\mu_n)}\right) \leq 2e^{-\frac{C^2\mu_n\log(\mu_n)}{2\left(\mu_n + C\sqrt{\mu_n\log(\mu_n)}\right)}} = O\left(\mu_n^{-\frac{C^2}{2}}\right). \qquad (1.13)$$

Chapter 2

Perfect matchings and Hamilton cycles

In this chapter, we will prove Theorem 1.3.1 claiming that in the regime $\alpha < \frac{1}{2}$, a.a.s., the existence of a Hamilton cycle as well as of a (near) perfect matching has a non-trivial phase transition in ν. In a nutshell, the proof of Theorem 1.3.1 has two parts: in the first part, for ν small enough, we count two collections of vertices: firstly the vertices close to the boundary of the disk having no neighbour close to the boundary of the disk and secondly, the vertices relatively close to the centre of the disk. Hence, all vertices of the first type have to be matched to vertices of the second type, but for ν small enough, there are more vertices of the first type than of the second type. For the second part, for ν large enough, we give an algorithm that maintains a set of vertex-disjoint cycles and isolated vertices. It iterates over the tiles of a suitable tessellation of the disk. In each step, it merges previous cycles or isolated vertices, s.t. upon termination it ends with a (single) Hamilton cycle. The fact that ν is large enough makes the density of vertices in each cell of the tessellation high enough so that this procedure terminates successfully.

2.1 Non-existence of perfect matching for sufficiently small ν

The following theorem yields the first part of Theorem 1.3.1.

Theorem 2.1.1. *For all positive real $\alpha < \frac{1}{2}$, there is a $\nu_0 = \nu_0(\alpha) > 0$ such that for all $0 < \nu < \nu_0$, the random graph $G(n; \alpha, \nu)$ does not have a near perfect matching a.a.s.*

Proof. The strategy is as follows. Let $s = \frac{1}{\alpha} > 2$. Let N_s be the number of vertices with radial coordinate at least $R - s$ and with no neighbour with radial coordinate at least $R - s$. Let M_s be the number of vertices with radial coordinate at most $R - s$. Hence, M_s is the number of vertices of $G(n; \alpha, \nu)$ inside the disk of radius

$R-s$ and N_s is a subset of the annulus $\mathcal{A}_s = \mathcal{D}_R \setminus \mathcal{D}_{R-s}$ of width s. If there is a perfect matching, then $M_s \geq N_s$ because a vertex with no neighbour with radius at least $R-s$ must be matched to a vertex with radius less than $R-s$, so distinct vertices counted by N_s must be matched to distinct vertices counted by M_s.

We observe that M_s is a binomial random variable with parameters n and 'success probability'

$$\frac{\cosh \alpha(R-s) - 1}{\cosh \alpha R - 1} \sim \frac{\frac{1}{2}e^{\alpha(R-s)}}{\frac{1}{2}e^{\alpha R}} = e^{-\alpha s} = e^{-1}.$$

(Here and elsewhere we write $a_n \sim b_n$ to denote that $a_n/b_n = 1 + o(1)$.)

Therefore its expectation is $\mathbb{E}M_s \sim ne^{-1}$ and its variance is $Var(M_s) \sim ne^{-1}(1-e^{-1})$. By Chebychev, it follows that for all $\epsilon > 0$,

$$\mathbb{P}(|M_s - \mathbb{E}M_s| \geq \epsilon \mathbb{E}M_s) \leq \frac{Var(M_s)}{\epsilon^2(\mathbb{E}M_s)^2} \sim \frac{e(1-e^{-1})}{\epsilon^2 n} = o(1).$$

(or in other words, $M_s \sim \mathbb{E}M_s$ a.a.s.)

Our aim now is to give a lower bound on $\mathbb{E}N_s$ and to show that $Var(N_s) = o((\mathbb{E}N_s)^2)$ as $n \to \infty$. For this, we label the vertex set $V = \{v_1, \ldots, v_n\}$, where v_i has radial coordinate r_i and write D_i^- for the number of neighbours of v_i with radial coordinate $\geq R-s$ and define the event

$$A_i = \{D_i^- = 0, r_i \geq R-s\}.$$

Then,

$$N_s = \sum_{i=1}^{n} \mathbb{1}_{A_i}.$$

As the vertices are identically distributed, we infer from this that

$$\mathbb{E}N_s = \sum_{i=1}^{n} \mathbb{P}(A_i) = n\mathbb{P}(A_1).$$

We condition on the radius $r_1 \in [R-s, R]$ which fully determines the geometric neighbourhood ball: if we know the radius r_1 of the vertex, then its number D_1^- of neighbours with radius at least $R-s$ is a binomial random variable with parameters $N-1$ and 'success probability' $p_1(r_1)$ which is given by an integral of the density function of the (α, R)-hyperbolic uniform distribution over the subset of the disk

$$\{(r,\theta) \in \mathbb{R}^2 : R-s \leq r < R, |\theta| \leq \vartheta(r, r_1)\},$$

i.e.

$$p_1(r_1) = \int_{R-s}^{R} \frac{2\vartheta(r,r_1)}{2\pi} \frac{\alpha \sinh \alpha r}{\cosh \alpha R - 1} dr.$$

From Lemma 1.6.3, we infer for R large enough and $r, r_1 \geq R - s$,

$$\vartheta(r, r_1) \leq 2e^{\frac{R-r-r_1}{2}} + Ke^{\frac{3}{2}(R-r-r_1)} \leq 3e^{s-\frac{R}{2}} =: c_\alpha \frac{\nu}{n}.$$

This implies that $p_1(r_1) \leq c_\alpha \frac{\nu}{n}$.

As D_1^- given r_1 is binomial, $\mathbb{P}(D_1^- = 0|r_1) = (1-p_1(r_1))^{n-1} \geq (1-\frac{\nu}{n}c_\alpha)^{n-1} = (1+o(1))e^{-\nu c_\alpha}$.

So, the conditioning on the radius r_1 leads to

$$\begin{aligned}
\mathbb{E}N_s &= n\int_{R-s}^{R} \mathbb{P}(D_1^- = 0|r_1)\frac{\alpha \sinh \alpha r_1}{\cosh \alpha R - 1}dr_1 \\
&\geq (1+o(1))ne^{-\nu c_\alpha}\int_{R-s}^{R} \frac{\alpha \sinh \alpha r_1}{\cosh \alpha R - 1}dr_1 \\
&= (1+o(1))ne^{-\nu c_\alpha}\int_{R-s}^{R} \alpha e^{\alpha r_1 - \alpha R}dr_1 \\
&= (1+o(1))ne^{-\alpha R}e^{-\nu c_\alpha}(e^{\alpha R} - e^{\alpha R - \alpha s}) \\
&= (1+o(1))ne^{-\nu c_\alpha}(1-e^{-\alpha s}) = (1+o(1))ne^{-\nu c_\alpha}(1-e^{-1}).
\end{aligned}$$

To show the concentration of N_s, observe that due to linearity of expectation and that the X_i and r_i are identically distributed,

$$\begin{aligned}
\mathbb{E}N_s^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{P}(A_i \cap A_j) = \sum_{i=1}^{n}\mathbb{P}(A_1) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{P}(A_1 \cap A_2) \\
&= n\mathbb{P}(A_1) + n(n-1)\mathbb{P}(A_1 \cap A_2) = \mathbb{E}N_s + n(n-1)\mathbb{P}(A_1 \cap A_2). \qquad (2.1)
\end{aligned}$$

Consider the event E that the angular distance between v_1 and v_2 is $\geq 4e^{\frac{2s-R}{2}} + 2Ke^{\frac{3}{2}(2s-R)}$ with K as from Lemma 1.6.3. If E holds, then there is no vertex v_0 with radial coordinate $r_0 \geq R-s$ which is adjacent to both v_1 and v_2 in the event $A_1 \cap A_2$ where v_1 and v_2 have radial coordinate $\geq R-s$. If there was, then the angular distance between v_1 and v_2 would be upper bounded by the sum of the upper bounds from Lemma 1.6.3, which is

$$2e^{\frac{R-r_1-r_0}{2}} + Ke^{\frac{3}{2}(R-r_1-r_0)} + 2e^{\frac{R-r_0-r_2}{2}} + Ke^{\frac{3}{2}(R-r_0-r_2)} \leq 4e^{\frac{2s-R}{2}} + 2Ke^{\frac{3}{2}(2s-R)}.$$

Therefore,

$$\mathbb{P}(E^c) \leq \frac{1}{2\pi}(4e^{\frac{2s-R}{2}} + 2Ke^{\frac{3}{2}(2s-R)}) = \frac{1}{2\pi}\left(4e^s\frac{\nu}{n} + 2Ke^{3s}\frac{\nu^3}{n^3}\right) = O\left(\frac{1}{n}\right), \qquad (2.2)$$

and in particular, $\mathbb{P}(E) > 0$.

We will prove the following auxiliary statement:

$$\mathbb{P}(A_1 \cap A_2) \leq \mathbb{P}(A_1)^2 + O\left(\frac{\mathbb{P}(A_1)}{n}\right). \qquad (2.3)$$

Let $B_1^- = \{p \in \mathcal{A}_s : d_{\mathbb{H}}(v_1, p) < R\}$ (where $d_{\mathbb{H}}$ is the hyperbolic distance) and $B_2^- = \{p \in \mathcal{A}_s : d_{\mathbb{H}}(v_2, p) < R\}$ denote the subset of those points of the disk which have radial coordinate $\geq R - s$ and would be adjacent to v_1 and v_2 respectively. Write $\mathbb{P}(v_i \in B_1^- | r_1) = p_1(r_1)$ and $\mathbb{P}(v_i \in B_2^- | r_2) = p_2(r_2)$ (note that this probability is indeed the same for all $i = 3, \dots, n$ because $v_3, \dots, v_n$ are i.i.d.). By conditioning on the radial coordinate of v_1 and v_2 we have:

$$\mathbb{P}(A_1 \cap A_2) = \int_{R-s}^{R} \int_{R-s}^{R} \mathbb{P}(A_1 \cap A_2 | r_1, r_2) \frac{\alpha^2 \sinh \alpha r_1 \sinh \alpha r_2}{(\cosh \alpha R - 1)^2} dr_1 dr_2. \tag{2.4}$$

If we condition further that the event E holds, then B_1^- and B_2^- are disjoint and hence for $i = 3, \dots, n$,

$$\begin{aligned} \mathbb{P}(v_i \notin B_1^- \cup B_2^- | r_1, r_2, E) &= 1 - p_1(r_1) - p_2(r_2) \\ &\leq 1 - p_1(r_1) - p_2(r_2) + p_1(r_1)p_2(r_2) \\ &= (1 - p_1(r_1))(1 - p_2(r_2)). \end{aligned}$$

If E does not hold, then for $i = 3, \dots, n$, we have $\mathbb{P}(v_i \notin B_1^- \cup B_2^- | r_1, r_2, E^c) \leq 1 - p_1(r_1)$. We infer that

$$\begin{aligned} &\mathbb{P}(A_1 \cap A_2 | r_1, r_2) \\ &= \mathbb{P}(A_1 \cap A_2 | r_1, r_2, E)\mathbb{P}(E | r_1, r_2) + \mathbb{P}(A_1 \cap A_2 | r_1, r_2, E^c)\mathbb{P}(E^c | r_1, r_2) \\ &= \mathbb{P}(A_1 \cap A_2 | r_1, r_2, E)\mathbb{P}(E) + \mathbb{P}(A_1 \cap A_2 | r_1, r_2, E^c)\mathbb{P}(E^c) \\ &\leq \prod_{i=3}^{n} \mathbb{P}(v_i \notin B_1^- \cup B_2^- | r_1, r_2, E) + \prod_{i=3}^{n} \mathbb{P}(v_i \notin B_1^- \cup B_2^- | r_1, r_2, E^c)\mathbb{P}(E^c) \\ &\leq \mathbb{P}(A_1 | r_1)\mathbb{P}(A_2 | r_2) + \mathbb{P}(A_1 | r_1) O\left(\frac{1}{n}\right). \end{aligned}$$

With this we conclude from (2.4),

$$\mathbb{P}(A_1 \cap A_2) \leq \mathbb{P}(A_1)\mathbb{P}(A_2) + \mathbb{P}(A_1) O\left(\frac{1}{n}\right),$$

which is the auxiliary claim as $\mathbb{P}(A_1) = \mathbb{P}(A_2)$.

Continuing from (2.1) and using the auxiliary claim (2.3),

$$\begin{aligned} \mathbb{E}N_s^2 = \mathbb{E}N_s + n(n-1)\mathbb{P}(A_1 \cap A_2) &\leq \mathbb{E}N_s + n^2 \left(\mathbb{P}(A_1)^2 + O\left(\frac{\mathbb{P}(A_1)}{n}\right)\right) \\ &= \mathbb{E}N_s + (\mathbb{E}N_s)^2 + O(n\mathbb{P}(A_1)) = (\mathbb{E}N_s)^2 + O(\mathbb{E}N_s). \end{aligned}$$

Therefore, the variance

$$Var(N_s) = \mathbb{E}N_s^2 - (\mathbb{E}N_s)^2 = O(\mathbb{E}N_s) = o((\mathbb{E}N_s)^2).$$

By Chebychev, it follows that for all $\epsilon > 0$,

$$\mathbb{P}(|N_s - \mathbb{E}N_s| \geq \epsilon \mathbb{E}N_s) \leq \frac{Var(N_s)}{\epsilon^2(\mathbb{E}N_s)^2} = o(1).$$

(Or in other words, $N_s \sim \mathbb{E}N_s$ a.a.s.)

In total, we see that $\mathbb{E}M_s \sim e^{-1}n$ and $\mathbb{E}N_s \geq (1+o(1))e^{-\nu c_\alpha}(1-e^{-1})n$ are both linear in n asymptotically. Due to the concentration, we have that

$$\mathbb{P}(M_s \leq (1+\epsilon)e^{-1}n) = 1 - o(1),$$

and

$$\mathbb{P}(N_s \geq (1-\epsilon)e^{-\nu c_\alpha}(1-e^{-1})n) = 1 - o(1).$$

By choosing ϵ and ν small enough, it follows that

$$\mathbb{P}(M_s < N_s) = 1 - o(1),$$

i.e. for ν sufficiently small, there is no near perfect matching a.a.s. □

2.2 Existence of Hamilton cycles for sufficiently large ν

The aim of this section is to prove the existence of a Hamilton cycle in $G(n;\alpha,\nu)$ with sufficiently high probability if ν is large enough.

The existence of a Hamilton cycle or the existence of a (near) perfect matching are examples of graph properties (recall that a graph property is a family of graphs closed under isomorphism). So, any result on a graph property can be applied to both the existence of a (near) perfect matching and the existence of a Hamilton cycle. The following observation is well known. We include its proof here for completeness.

Lemma 2.2.1. *If $G_{Po}(n;\alpha,\nu)$ denotes the Poissonized KPKVB model and $\mathcal{V}_{\mathrm{Po}}$ its vertex set, then*

$$\mathbb{P}(G_{Po}(n;\alpha,\nu) \in \mathcal{F} \mid |\mathcal{V}_{\mathrm{Po}}| = n) = 1 - o(1), \text{ if } \mathbb{P}(G_{Po}(n;\alpha,\nu) \in \mathcal{F}) = 1 - o(n^{-1/2}).$$

Thus, if $\mathbb{P}(G_{Po}(n;\alpha,\nu) \notin \mathcal{F}) = o(n^{-1/2})$, then $\mathbb{P}(G(n;\alpha,\nu) \notin \mathcal{F}) = o(1)$.

Proof. By Stirling's formula

$$\mathbb{P}(|\mathcal{V}_{\mathrm{Po}}| = n) = \frac{n^n}{n!}e^{-n} = (1+o(1))\frac{n^n}{\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}}e^{-n} = (1+o(1))\frac{1}{\sqrt{2\pi n}}.$$

So as $n \to \infty$, writing $E_n := \{G_{\mathrm{Po}}(n;\alpha,\nu) \in \mathcal{F}\}$,

$$\mathbb{P}(E_n^c) \geq \mathbb{P}(E_n^c \mid |\mathcal{V}_{\mathrm{Po}}| = n) \cdot \Theta\left(n^{-1/2}\right).$$

Therefore, if $\mathbb{P}(E_n^c) = o(n^{-1/2})$, we deduce that $\mathbb{P}(E_n^c \mid |\mathcal{V}_{\mathrm{Po}}| = n) = o(1)$. □

We will apply a standard depoissonisation technique: using Lemma 2.2.1, we will develop our arguments in the poissonisation of the KPKVB model. More precisely, we will show that $G_{\mathrm{Po}}(n;\alpha,\nu)$ satisfies certain events with sufficiently high probability (that is, with probability at least $1-o(n^{-1/2})$), and we then use Lemma 2.2.1 to deduce that $G(n;\alpha,\nu)$ also satisfies them a.a.s., that is, with probability $1-o(1)$.

Theorem 2.2.2. *For all positive real $\alpha < \frac{1}{2}$, there is a $\nu_1 = \nu_1(\alpha)$ such that for all $\nu > \nu_1$, the random graph $G_{Po}(n;\alpha,\nu)$ has a Hamilton cycle and hence also a near perfect matching with probability $1-o(n^{-1/2})$.*

2.2.1 A useful tiling

We consider the following tiling:

$$T_{i,j} = \{(r,\theta) \in \mathcal{D}_R : R-(i+1)2\ln 2 \leq r < R - i2\ln 2, j\frac{2\pi}{n_i} < \theta \leq (j+1)\frac{2\pi}{n_i}\},$$

where $n_i = n_{i,R} = 2^{4-i+\lfloor \frac{R}{2\ln 2}\rfloor} \in \mathbb{N}_0 := \mathbb{N}\cup\{0\}$, for $i \in \mathbb{N}_0$, $i \leq i_{max} = \lceil \frac{0.9R}{2\ln 2}\rceil$ and $j \in \mathbb{N}_0$, $j < n_i$. We call i,j *admissible* if they satisfy these constraints.

Note that the parameter n_i is an integer and, in fact, a power of 2, as the exponent $4-i+\lfloor\frac{R}{2\ln 2}\rfloor$ is a positive integer. Indeed, for $i \leq i_{max}$ the exponent is at least $4-\frac{0.9R}{2\ln 2}+\frac{R}{2\ln 2}-1 = 3+0.1\frac{R}{2\ln 2} > 0$.

We call the collection of tiles with a fixed given i the *i-th layer*. These are the tiles in the i-th annulus where we start counting from zero at the boundary of the disk. Note that there are n_i tiles in the i-th layer and the tiling covers the annulus with exterior radius R and interior radius $R - i_{max}2\ln 2 = (1+o(1))0.1R$ (in particular, the most interior layer i_{max} is contained in the smaller disk with radius $\frac{R}{2}$ around the origin). A schematic picture is shown in Figure 2.1.

We say that a tile $T_{i',j'}$ is *below* the tile $T_{i,j}$, if $i' \leq i$ and the smallest angular sector containing $T_{i,j}$ also contains $T_{i',j'}$.

Lemma 2.2.3 (Adjacency among the tiles)**.** *For admissible indices i,j, any point $p \in T_{i,j}$ is within distance R from any point p' in any tile below the tile $T_{i,j}$.*

Proof. Let $p = (r,\theta) \in T_{i,j}$ and $p' = (r',\theta') \in T_{i',j'}$ be a vertex in any tile below tile $T_{i,j}$ (in the sense of the statement above). Note that $r' \geq r$ must hold. Then, the angular distance $|\theta-\theta'|_{2\pi}$ between p and p' is at most the angular width of the tile $T_{i,j}$, which is

$$\frac{2\pi}{n_i} = 2^{-3}\pi 2^{i-\lfloor\frac{R}{2\ln 2}\rfloor} \leq 2^{i-1}e^{-\frac{R}{2}}.$$

On the other hand, we know that the radial coordinates satisfy $r \leq R - i2\ln 2$ and $r' < R$. If $r+r' \leq R$, we have adjacency by the triangle inequality. If $r+r' > R$ and using that $r, r' \geq (1+o(1))0.1R$ (as remarked earlier), we distinguish two cases:

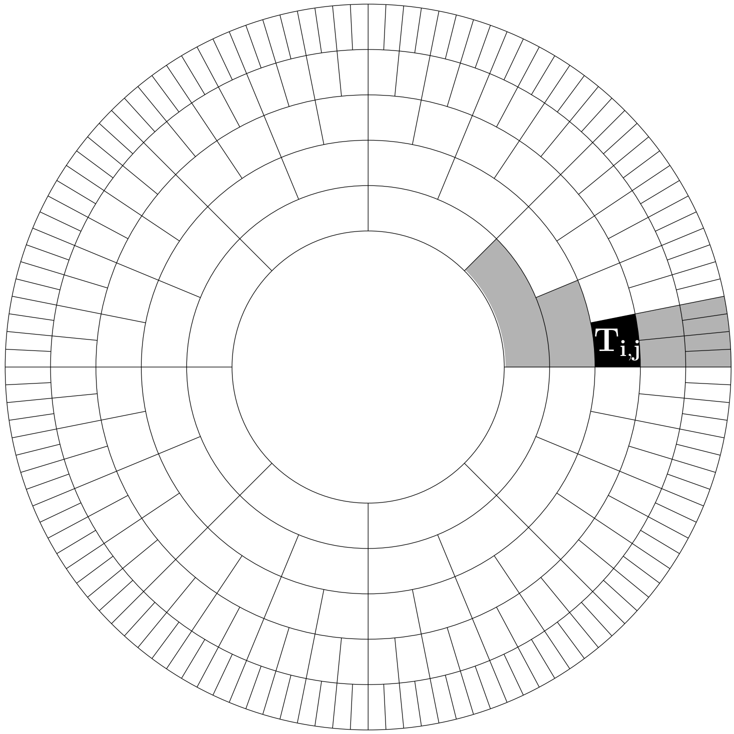

Figure 2.1: (Partial) tiling in the hyperbolic disk; example of a tile $T_{i,j}$ (coloured black) and the tiles which are guaranteed to lie within its neighbourhood ball by Lemma 2.2.3 (coloured black and grey).

1. If $r, r' < R - K$ (with K as in Lemma 1.6.3), then by the last part of Lemma 1.6.3, it holds that

$$\vartheta(r, r') \geq 2e^{\frac{R-r-r'}{2}} \geq 2e^{\frac{i2\ln 2 - R}{2}} = 2^{i+1}e^{-\frac{R}{2}}.$$

2. Otherwise we may assume that $r' \geq R - K$ holds, while still $r, r' \geq (1 + o(1))0.1R$. Therefore, $R - r - r' \leq R - (R - K) - (1 + o(1))0.1R = -(1 + o(1))0.1R + K$, hence the error term in Lemma 1.6.3 is $Ke^{\frac{3}{2}(R-r-r')} = o(e^{\frac{1}{2}(R-r-r')})$ and it follows that

$$\vartheta(r, r') \geq 2e^{\frac{R-r-r'}{2}} - o(e^{\frac{1}{2}(R-r-r')}) > e^{\frac{R-r-r'}{2}} \geq e^{\frac{i2\ln 2 - R}{2}} = 2^i e^{-\frac{R}{2}}.$$

We conclude that $|\theta - \theta'|_{2\pi} \leq \vartheta(r, r')$ from which the claim follows. $\square$

We will denote the number of points falling into $T_{i,j}$ by $N(T_{i,j})$. Note that for the Poissonized KPKVB model G_{Po}, the family of random variables $N(T_{i,j})$ for all admissible i, j is independent.

Lemma 2.2.4 (Expected number of points in a tile)**.** *Let $\alpha, \nu > 0$. For admissible indices i, j, the expected number of points falling into $T_{i,j}$ satisfies*

$$\mathbb{E}[N(T_{i,j})] = \Theta(\nu 2^{i(1-2\alpha)}).$$

Proof. The expected number of points falling into $T_{i,j}$ is given by

$$\begin{aligned}
\mu_g(T_{i,j}) &= n \cdot \int_{R-(i+1)2\ln 2}^{R-i2\ln 2} \int_{j\frac{2\pi}{n_i}}^{(j+1)\frac{2\pi}{n_i}} \frac{\alpha \sinh \alpha r}{2\pi(\cosh \alpha R - 1)} d\theta dr \\
&= n \cdot \int_{R-(i+1)2\ln 2}^{R-i2\ln 2} \frac{\alpha \sinh \alpha r}{n_i(\cosh \alpha R - 1)} dr \\
&= n \cdot \frac{\cosh(\alpha(R - i2\ln 2)) - \cosh(\alpha(R-(i+1)2\ln 2))}{n_i(\cosh \alpha R - 1)}.
\end{aligned}$$

As $i \leq i_{max}$, we have that $R - i2\ln 2 \geq 0.1R \to \infty$, and hence

$$\cosh(\alpha(R - i2\ln 2)) = (1+o(1))\frac{1}{2}e^{\alpha(R-i2\ln 2)},$$

$$\cosh(\alpha(R - (i+1)2\ln 2)) = (1+o(1))\frac{1}{2}e^{\alpha(R-(i+1)2\ln 2)},$$

and

$$\cosh \alpha R = (1+o(1))\frac{1}{2}e^{\alpha R}.$$

Furthermore, $n_i = 2^{4-i+\lfloor \frac{R}{2\ln 2} \rfloor} = \Theta(2^{-i+\frac{R}{2\ln 2}}) = \Theta(2^{-i}e^{\frac{R}{2}})$. We conclude

$$\begin{aligned}
\mathbb{E}[N(T_{i,j})] &= \Theta\left(n \frac{e^{\alpha(R-i2\ln 2)} - e^{\alpha(R-(i+1)2\ln 2)}}{2^{-i}e^{\frac{R}{2}}e^{\alpha R}}\right) \\
&= \Theta\left(n 2^i e^{-\frac{R}{2}} e^{-i2\alpha \ln 2}(1 - e^{-2\alpha \ln 2})\right) \\
&= \Theta(n 2^{i(1-2\alpha)} e^{-\frac{R}{2}}).
\end{aligned}$$

Finally, using that $R = 2\ln\frac{n}{\nu}$, that is, $n = \nu e^{\frac{R}{2}}$, yields the claim. □

2.2.2 A procedure for finding a Hamilton cycle

In this subsection we describe the strategy of our procedure for finding a Hamilton cycle in a graph which is embedded in the hyperbolic disk $\mathcal{D}_R$ and which makes use of the tiling $(T_{i,j})_{i,j\in\mathbb{N}_0, i\leq i_{max}, j<n_i}$ defined above. Roughly speaking, the procedure iterates through the layers of the tiling, working upwards from the 0-th layer to layer $i_{\max}$, gathering a suitable collection of vertex-disjoint cycles and isolated vertices. When processing the tile $T_{i,j}$, it merges as many vertex-disjoint cycles and isolated vertices from previous iterations that are below the tile as possible. Once the procedure has reached the maximum layer which is completely

contained in the smaller disk with radius $\frac{R}{2}$, the procedure attempts to merge all the remaining cycles and vertices.

We now describe the procedure in more detail. For each tile $T_{i,j}$ we will define a random variable $D_{i,j}$ called demand, which will be used later in the probabilistic analysis to show that the procedure terminates successfully. Recall that $N(T_{i,j})$ denotes the number of vertices in tile $T_{i,j}$ and note that the collection of $N(T_{i,j})$ for admissible i, j are independent Poisson random variables for the poissonised KPKVB model.

We say that a tile $T_{i,j}$ can be *covered by x vertex-disjoint cycles and isolated vertices* if the set of vertices in tile $T_{i,j}$ can be partitioned into x sets and each set that is not a single vertex constitutes a cycle (with at least 3 vertices), where we ignore additional edges, e.g. such a set in the partition could even be a clique.

Lemma 2.2.5 (Cycle merging, see Figure 2.2)**.** *If the vertices strictly below tile $T_{i,j}$ can be covered by x vertex-disjoint cycles and isolated vertices and the number $y = N(T_{i,j})$ of vertices in tile $T_{i,j}$ satisfies $y \geq 3$, then the set of all vertices below $T_{i,j}$ (including those in $T_{i,j}$) can be covered by* $\max\{1, x-y+1\}$ *cycles and isolated vertices.*

Furthermore, if additionally $y > x$, then the vertices below $T_{i,j}$ can be covered by a single cycle which has $y - x$ edges within $T_{i,j}$.

Proof. If $y = N(T_{i,j}) \geq 3$, then the vertices in $T_{i,j}$ form a cycle by Lemma 2.2.3 (in fact from that lemma we know that they even form a clique). Each of the y edges of this cycle in $T_{i,j}$ can be used to merge it with a cycle or vertex strictly below $T_{i,j}$: to merge the cycle in tile $T_{i,j}$ with a cycle strictly below $T_{i,j}$, use an edge $e_i = v_i v_{i+1}$ of the cycle $v_1, \ldots, v_y$ in $T_{i,j}$ and choose any edge $e_* = a_* b_*$ from the cycle below. By Lemma 2.2.3, the four endpoints form a clique and therefore, we can go along the edges $v_i a_*$, then the cycle below (without the edge e_*), and finally along $b_* v_{i+1}$ to bring us back to the cycle in $T_{i,j}$. To merge the cycle in tile $T_{i,j}$ with a vertex a_* below $T_{i,j}$, we can just use the edges $v_i a_*$ and $a_* v_{i+1}$ instead of $v_i v_{i+1}$.

If $y \leq x$, then all edges of the original cycle in $T_{i,j}$ will be used and we end up with $x - y + 1 \geq 1$ cycles and (isolated) vertices below (and including) $T_{i,j}$. If $y > x$, then all cycles and (isolated) vertices strictly below $T_{i,j}$ become part of the original cycle in $T_{i,j}$ and $y - x > 0$ edges of the cycle in $T_{i,j}$ remain unused and part of the final cycle. $\square$

The demand random variables $D_{i,j}$ for admissible i, j are defined in terms of the point counts $N(T_{i,j})$ as follows. For $i = 0$ and $j = 0, \ldots, n_0 - 1$ we set:

$$D_{0,j} = \begin{cases} N(T_{i,j}), & \text{if } N(T_{i,j}) \in \{1, 2\}, \\ 0, & \text{otherwise,} \end{cases}$$

and, for $0 < i \leq i_{\max}$ and $j = 0, \ldots, n_i - 1$ we set:

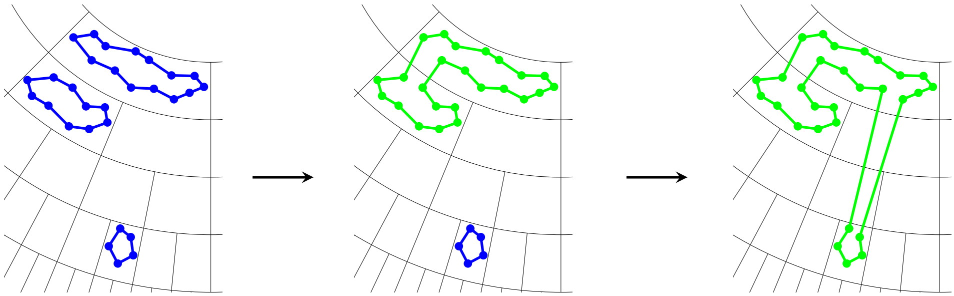

Figure 2.2: In three steps, three cycles (coloured blue) are merged (resulting in the green cycle) by replacing an edge of one cycle by a detour around the other cycle. Note that we zoomed into the part of the disk which matters for the cycle merging.

$$D_{i,j} = \max\{D_{i-1,2j} + D_{i-1,2j+1} + 3 - N(T_{i,j}), 0\}.$$

In particular $D_{i,j}$ and $D_{i,j'}$ are independent for $j \neq j'$, since they depend on disjoint regions. Also, the $D_{0,j}$ are i.i.d. random variables with values in $\{0, 1, 2\}$ satisfying

$$\mathbb{P}(D_{0,j} = 1) = \mu_0 e^{-\mu_0}, \qquad \mathbb{P}(D_{0,j} = 2) = \frac{\mu_0^2}{2} e^{-\mu_0},$$

where we have used the notation $\mu_0 := \mu_g(T_{0,0}) \overset{\text{Lemma 2.2.4}}{=} \Theta(\nu)$.

Lemma 2.2.6. *For admissible indices i, j, if $D_{i,j} = x$, then the vertices below (and in) $T_{i,j}$ can be covered by at most $x + 1$ vertex-disjoint cycles and isolated vertices (in total).*

Moreover, if $i > 0$ and $D_{i,j} = 0$, then the vertices below (and in) $T_{i,j}$ can be covered by exactly one cycle which has at least one edge which is completely contained in $T_{i,j}$.

Proof. The proof is by induction on i. For $i = 0$, the claim is clear because then $D_{0,j} = 0$ implies that there is either one cycle or no vertex in $T_{i,j}$.

For $i > 0$, assuming the claim for $i - 1$ we show it for i. By the induction hypothesis, the vertices below $T_{i-1,2j}$ ($T_{i-1,2j+1}$, respectively) can be covered by $D_{i-1,2j} + 1$ ($D_{i-1,2j+1} + 1$, respectively) many vertex-disjoint cycles and isolated vertices. Thus, the vertices strictly below $T_{i,j}$ can be covered by $D_{i-1,2j} + D_{i-1,2j+1} + 2$ many vertex-disjoint cycles and isolated vertices in total. If $N(T_{i,j}) \geq 3$, then by Lemma 2.2.5, the vertices below $T_{i,j}$ can be covered by

$$\begin{aligned} &\max\{1, D_{i-1,2j} + D_{i-1,2j+1} + 2 - N(T_{i,j}) + 1\} \\ &\leq \max\{D_{i-1,2j} + D_{i-1,2j+1} + 3 - N(T_{i,j}), 0\} + 1 \end{aligned}$$

$$= D_{i,j} + 1.$$

If $N(T_{i,j}) \le 2 < 3$, then the vertices in $T_{i,j}$ just remain as isolated vertices and the area in and below $T_{i,j}$ can be covered by at most $D_{i-1,2j} + D_{i-1,2j+1} + 4 \le D_{i,j} + 1$ vertex-disjoint cycles and isolated vertices.

In particular, if $D_{i,j} = 0$, the vertices in the area in and below $T_{i,j}$ can be covered by one cycle or vertex. If $i > 0$, then the condition $D_{i,j} = 0$ and the definition of $D_{i,j}$ imply that there are at least 3 vertices in $T_{i,j}$. Hence, the vertices in and below $T_{i,j}$ can be covered by exactly one cycle, which will have at least one edge with both endpoints inside $T_{i,j}$ (using that $N(T_{i,j}) > D_{i-1,2j} + D_{i-1,2j+1} + 2$). □

Lemma 2.2.7. *If $D_{i,j} = 0$ for $i = i_{max}$ and for all $j = 0, \dots, n_i - 1$, then there is a Hamilton cycle.*

Proof. Firstly, we observe that for $i = i_{max}$, if $D_{i,j} = 0$, then all vertices in and below $T_{i,j}$ can be covered by one cycle that contains an edge whose endpoints are both in $T_{i,j}$ by Lemma 2.2.6. Taking such an edge for $T_{i,0}$ and $T_{i,1}$, the four endpoints form a clique by the triangle inequality because all radial coordinates are at most $\frac{R}{2}$ and hence the cycle of $T_{i,1}$ can be taken as a detour to the cycle of $T_{i,0}$ as in the proof of Lemma 2.2.5. As a result, we have a cycle covering all vertices below $T_{i,0}$ and $T_{i,1}$ and with an edge inside the i-th layer. We can repeat this procedure to merge this resulting cycle also with those in $T_{i,2}, \dots, T_{i,n_i-1}$. We will end up with one cycle covering all vertices below all tiles $T_{i,0}, \dots, T_{i,n_i-1}$, and this cycle contains an edge whose endpoints are both in the inner disk with radius $\frac{R}{2}$. The remaining vertices in the inner disk, that are not in any tile, form a clique and in particular can be covered by a cycle. We can again merge this cycle with the one we have created earlier via the same trick. □

2.2.3 Probabilistic lemmas which ensure the a.a.s. successful termination of the procedure

In this subsection we show that the algorithm explained previously works successfully for the Poissonised KPKVB model $G_{\mathrm{Po}}(n; \alpha, \nu)$ with $N \overset{d}{=} \mathrm{Po}(n)$ many vertices (the standard depoissonisation of Lemma 2.2.1 gives then the result in the standard KPKVB model). Lemma 2.2.9 of this section shows the exponential decay of the demand random variables, which we then use in Lemma 2.2.10 to show that the demand random variables are simultaneously zero in the maximum layer. Appealing to Lemma 2.2.7, we can then conclude that this makes the algorithm work.

Sub-exponential tail decay of demand

We first show the following technical lemma:

Lemma 2.2.8. *For all real $\epsilon \in (0,1)$, there exists $\kappa = \kappa(\epsilon) > 0$ such that for all $i \in \mathbb{N}_{>0}$, for all $x \geq \kappa i^2 \ln(1+i)$ we have: $(x+1)e^{-\frac{x}{i^2}} \leq \epsilon$.*

Proof. Pick $\kappa > \max\{\frac{1}{\ln 2}, 3\}$ such that $(\kappa+1)2^{3-\kappa} \leq \epsilon$; this is possible as $\lim_{a\to\infty}(a+1)2^{3-a} = 0$. We prove the lemma in the following way: in the first step we verify that for $x = \kappa i^2 \ln(1+i)$ we have $(x+1)e^{-\frac{x}{i^2}} \leq \epsilon$, and then we show that the left-hand side of the inequality is monotone decreasing in x (by showing that its derivative with respect to x is negative). Since the right-hand side is independent of x, this clearly implies the lemma.

For the first step, we need to show that $(\kappa i^2 \ln(1+i)+1)e^{-\kappa \ln(1+i)} \leq \epsilon$. Using that $i^2 \leq (1+i)^2$, $\ln(1+i) \leq 1+i$ and $1 \leq (1+i)^3$, we note that the left-hand side of the inequality can be bounded from above by

$$(\kappa i^2 \ln(1+i)+1)e^{-\kappa \ln(1+i)} \leq (\kappa+1)(1+i)^3(1+i)^{-\kappa} = (\kappa+1)(1+i)^{3-\kappa}.$$

Now, if we plug in $i = 1$, this upper bound is at most ϵ by the choice of κ. The derivative in i of the latter expression is

$$(\kappa+1)(3-\kappa)(1+i)^{2-\kappa},$$

which is negative for $\kappa > 3$ for all $i \geq 1$, Therefore, the upper bound is monotone decreasing in i and hence, for all $i \geq 1$ and $x = \kappa i^2 \ln(1+i)$, we have $(x+1)e^{-\frac{x}{i^2}} \leq \epsilon$, concluding the first step.

For the second step, we need to verify that the derivative of $(x+1)e^{-\frac{x}{i^2}}$ is negative in $x \geq \kappa i^2 \ln(1+i) > 0$: using the assumptions of $\kappa > \frac{1}{\ln 2}$ and $i \geq 1$, we have

$$e^{-\frac{x}{i^2}}\left(1+(x+1)\left(-\frac{1}{i^2}\right)\right) \leq 1 - \kappa \ln(1+i) - \frac{1}{i^2} \leq 1 - \kappa \ln 2 < 0.$$

The lemma follows. □

We are now ready to state and prove the main lemma of this section.

Lemma 2.2.9. *There is a constant $c > 0$ such that for $0 < \alpha < \frac{1}{2}$ and ν sufficiently large, for all admissible i, j, and all $t \geq 0$:*

$$\mathbb{P}(D_{i,j} \geq t) \leq e^{-ct}.$$

Proof. Set $c = 10$, $c_0 = c + \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$ and $c_i = c_{i-1} - \frac{1}{i^2}$, for $i > 0$. So, in particular, we have $\infty > c_0 > c_1 > \cdots > c = 10 > 0$.

We prove the lemma by induction on i. For the base case $i = 0$, the claim is clear for $t > 2$ because $D_{0,j} \in \{0,1,2\}$, so $\mathbb{P}(D_{0,j} \geq t) = 0 < e^{-c_0 t}$. For $t = 1, 2$,

$$\mathbb{P}(D_{0,j} \geq t) \leq \mu_0 e^{-\mu_0} + \frac{\mu_0^2}{2} e^{-\mu_0} = O(\nu^2)e^{-\Theta(\nu)},$$

where the equality follows by Lemma 2.2.4. In particular, by choosing ν large enough, it holds that $\mathbb{P}(D_{0,j} \geq t) \leq e^{-c_0 t}$ for $t = 1, 2$.

For the inductive step, assume the statement is true for $i-1$ with $1 \leq i \leq i_{max}$. Note that as $D_{i-1,2j}$ and $D_{i-1,2j+1}$ are independent, we can apply the induction hypothesis to $D_{i-1,2j}$ and $D_{i-1,2j+1}$ to get

$$\begin{aligned}\mathbb{P}(D_{i-1,2j} + D_{i-1,2j+1} \geq t) &\leq \sum_{s=0}^{t} \mathbb{P}(D_{i-1,2j} \geq s)\mathbb{P}(D_{i-1,2j+1} \geq t-s) \\ &\leq \sum_{s=0}^{t} e^{-c_{i-1}s} e^{-c_{i-1}(t-s)} = (t+1)e^{-c_{i-1}t}. \qquad (2.5)\end{aligned}$$

Define

$$\epsilon = \min\left\{ e^{-3c_0}, \frac{1}{2}(1 - e^{-c}) \right\},$$

let $\kappa = \kappa(\epsilon)$ be as in Lemma 2.2.8 and set

$$t_i := \kappa i^2 \ln(1+i) + 3.$$

We make a case distinction depending on t.

Case 1: $t \geq t_i$.

Using the definition of $D_{i,j}$ and by (2.5), we have

$$\begin{aligned}\mathbb{P}(D_{i,j} \geq t) &= \sum_{s=0}^{\infty} \mathbb{P}(N(T_{i,j}) = s)\mathbb{P}(D_{i-1,2j} + D_{i-1,2j+1} \geq t+s-3) \\ &\overset{(2.5)}{\leq} \sum_{s=0}^{\infty} \mathbb{P}(N(T_{i,j}) = s)(t+s-3+1)e^{-c_{i-1}(t+s-3)}.\end{aligned}$$

Now, we can apply Lemma 2.2.8 to $x = t+s-3 \geq \kappa i^2 \ln(1+i)$ to deduce that

$$(x+1)e^{(c_i - c_{i-1})x} = (x+1)e^{-\frac{x}{i^2}} \leq \epsilon, \text{ which implies } (x+1)e^{-c_{i-1}x} \leq \epsilon e^{-c_i x}.$$

We infer that

$$\begin{aligned}\mathbb{P}(D_{i,j} \geq t) &\leq \sum_{s=0}^{\infty} \mathbb{P}(N(T_{i,j}) = s)\epsilon e^{-c_i(t+s-3)} \\ &= \epsilon e^{3c_i} e^{-c_i t} \sum_{s=0}^{\infty} \mathbb{P}(N(T_{i,j}) = s)e^{-c_i s} \\ &\leq e^{-c_i t} \sum_{s=0}^{\infty} \mathbb{P}(N(T_{i,j}) = s) = e^{-c_i t},\end{aligned}$$

where the third line follows by choice of ϵ and the definition of the sequence $c_0, c_1, \ldots$, which implies that $c_i < c_0$.

Case 2: $t < t_i$.

Let $\mu_i := \mu_{n,\alpha,\nu}(T_{i,0})$. We first observe that for all $i \in \mathbb{N}_0$:

$$\mu_i \geq (c_i t_i + \ln 2)\frac{2}{1 - \ln 2} \tag{2.6}$$

and

$$\frac{1}{2}\mu_i \geq t_i + 3. \tag{2.7}$$

To see that this holds, note that as $\mu_i = \Omega(\nu 2^{i(1-2\alpha)})$ (see Lemma 2.2.4), we can take any $\nu_* > 0$ and then pick $i_0 = i_0(\nu_*) \in \mathbb{N}$ such that for all $\nu \geq \nu_*$ and all $i \geq i_0$, the claims hold (as the right-hand side is independent of ν and grows at most polynomially in i whereas μ_i grows exponentially in i). Then, as the right-hand sides of (2.6) and (2.7) are independent of ν, we can pick $\nu_{**} > \nu_*$ large enough such that (2.6) and (2.7) also hold for $i = 0, \ldots, i_0(\nu_*)$.

Thus, (2.6) and (2.7) hold for all $\nu > \nu_{**}$ and all $i \in \mathbb{N}_0$.

We have

$$\begin{aligned}
&\mathbb{P}(D_{i,j} \geq t)\\
&= \sum_{j=0}^{\infty} \mathbb{P}(D_{i-1,2j} + D_{i-1,2j+1} = j) \times \\
&\qquad \mathbb{P}(D_{i-1,2j} + D_{i-1,2j+1} + 3 - N(T_{i,j}) \geq t | D_{i-1,2j} + D_{i-1,2j+1} = j)\\
&= \sum_{j=0}^{\infty} \mathbb{P}(D_{i-1,2j} + D_{i-1,2j+1} = j)\mathbb{P}(N(T_{i,j}) \leq j + 3 - t)\\
&= \sum_{j=\max\{t-3,0\}}^{\infty} \mathbb{P}(D_{i-1,2j} + D_{i-1,2j+1} = j)\mathbb{P}(N(T_{i,j}) \leq j + 3 - t)\\
&\leq \sum_{j=\max\{t-3,0\}}^{\infty} \mathbb{P}(D_{i-1,2j} + D_{i-1,2j+1} = j)\mathbb{P}(N(T_{i,j}) \leq j + 3).
\end{aligned}$$

We split the sum into two parts: for $j + 3 \leq \frac{1}{2}\mu_i$, we apply Lemma 1.6.6 with $k = j+3$ and $\mu = \mu_i$ and hence, $\frac{k}{\mu} = \frac{j+3}{\mu_i} \leq \frac{1}{2}$. Therefore, $H\left(\frac{k}{\mu}\right) \geq \frac{1}{2}(1-\ln 2) > 0$ and we get

$$\mathbb{P}(N(T_{i,j}) \leq j + 3) \leq e^{-\mu_i \frac{1}{2}(1-\ln 2)}.$$

By (2.6), it follows that

$$e^{-\mu_i \frac{1}{2}(1-\ln 2)} \leq \frac{1}{2}e^{-c_i t_i} \leq \frac{1}{2}e^{-c_i t}.$$

So, we have for the first part of the sum

$$\sum_{\substack{j=\max(t-3,0)\\ j+3\le\frac{1}{2}\mu_i}}^{\infty} \mathbb{P}(D_{i-1,2j}+D_{i-1,2j+1}=j)\mathbb{P}(N(T_{i,j})\le j+3) \le e^{-\mu_i\frac{1}{2}(1-\ln 2)}$$
$$\le \frac{1}{2}e^{-c_i t}.$$

For the second part, we have $j+3>\frac{1}{2}\mu_i$. By (2.7), $\frac{1}{2}\mu_i \ge t_i+3>t_i$. By (2.5) and Lemma 2.2.8 with $x=j\ge\kappa i^2\ln(1+i)$ it holds that

$$\begin{aligned}\mathbb{P}(D_{i-1,2j}+D_{i-1,2j+1}=j) &\le \mathbb{P}(D_{i-1,2j}+D_{i-1,2j+1}\ge j)\\ &\le (j+1)e^{-c_{i-1}j}\le \epsilon e^{-c_i j}.\end{aligned}$$

With this, we can also bound the second sum from above as

$$\begin{aligned}\sum_{\substack{j=\max\{t-3,0\}\\ j+3>\frac{1}{2}\mu_i}}^{\infty} \mathbb{P}(D_{i-1,2j}+D_{i-1,2j+1}=j)\mathbb{P}(N(T_{i,j})\le j+3) &\le \sum_{j=t}^{\infty}\epsilon e^{-c_i j}\\ &= e^{-c_i t}\epsilon\sum_{j=0}^{\infty}(e^{-c_i})^j\\ &= e^{-c_i t}\epsilon\frac{1}{1-e^{-c_i}}\\ &\le e^{-c_i t}\epsilon\frac{1}{1-e^{-c}}\le\frac{1}{2}e^{-c_i t}.\end{aligned}$$

where the last inequality follows from the choice of ϵ and the fact that $c_i>c$.

By combining both sums, we conclude that also for t as in Case 2, $\mathbb{P}(D_{i,j}\ge t)\le\frac{1}{2}e^{-c_i t}+\frac{1}{2}e^{-c_i t}=e^{-c_i t}$, and the lemma follows. $\square$

Deriving Theorem 2.2.2

Finally, Theorem 2.2.2 is a result of the following lemma together with Lemma 2.2.7:

Lemma 2.2.10. *Let $0<\alpha<\frac{1}{2}$, ν sufficiently large. Then*

$$\mathbb{P}(D_{i_{\max},j}=0,\ \textit{for all } j=0,\dots,n_{i_{\max}}-1)=1-o\left(n^{-1/2}\right).$$

Proof. First, let us recall that the number of tiles in the ith layer is $n_i=2^{4-i+\lfloor\frac{R}{2\ln 2}\rfloor}$. For $i=i_{max}=\lceil\frac{0.9R}{2\ln 2}\rceil$, it follows that $n_i=\Theta(2^{\frac{0.1R}{2\ln 2}})=\Theta(n^{0.1})$. Furthermore, $\mu_i=\Omega(2^{i_{max}(1-2\alpha)})=\Omega(n^{0.9(1-2\alpha)})$.

We have that

$$\mathbb{P}(\text{for all } j=0,\dots,n_i-1: D_{i,j}=0)=1-\mathbb{P}(D_{i,j}>0 \text{ for some } j).$$

By the union bound over all tiles in layer $i = i_{\max}$

$$\mathbb{P}(D_{i,j} > 0 \text{ for some } j) \leq \sum_{j=0}^{n_i-1} \mathbb{P}(D_{i,j} > 0).$$

Now we observe that if $D_{i-1,2j} \leq \frac{\mu_i}{100}$ and $D_{i-1,2j+1} \leq \frac{\mu_i}{100}$ and $N(T_{i,j}) \geq \frac{3}{100}\mu_i$ all hold, then $3 + D_{i-1,2j} + D_{i-1,2j+1} - N(T_{i,j}) \leq 3 + \frac{2}{100}\mu_i - \frac{3}{100}\mu_i \leq 0$ since $\mu_i = \Omega(\nu 2^{i(1-2\alpha)})$. Hence if all three of these conditions hold then $D_{i,j} \leq 0$. In other words, if $D_{i,j} > 0$ then $D_{i-1,2j} > \frac{\mu_i}{100}$ or $D_{i-1,2j+1} > \frac{\mu_i}{100}$ or $N(T_{i,j}) < \frac{3}{100}\mu_i$.

Therefore,

$$\begin{aligned}\mathbb{P}(D_{i,j} > 0) &\leq \mathbb{P}\left(D_{i-1,2j} > \frac{\mu_i}{100} \text{ or } D_{i-1,2j+1} > \frac{\mu_i}{100} \text{ or } N(T_{i,j}) < \frac{3\mu_i}{100}\right) \\ &\leq \mathbb{P}\left(D_{i-1,2j} > \frac{\mu_i}{100}\right) + \mathbb{P}\left(D_{i-1,2j+1} > \frac{\mu_i}{100}\right) + \mathbb{P}\left(N(T_{i,j}) < \frac{3\mu_i}{100}\right).\end{aligned}$$

For the first two terms, we use Lemma 2.2.9, taking ν sufficiently large, and for the third term, we apply Lemma 1.6.6. We get

$$\mathbb{P}(D_{i,j} > 0) \leq 2e^{-c\frac{\mu_i}{100}} + e^{-\Omega(\mu_i)} = e^{-\Omega(\mu_i)}.$$

Using that $\mu_i = \Omega(n^{0.9(1-2\alpha)})$, it follows that $\mathbb{P}(D_{i,j} > 0) = O(e^{-\Omega(n^{0.9(1-2\alpha)})})$. Since $n_{i_{max}} = \Theta(n^{0.1})$, we obtain

$$\mathbb{P}(D_{i,j} > 0 \text{ for some } j) \leq \sum_j \mathbb{P}(D_{i,j} > 0) = O\left(n^{0.1} e^{-\Omega(n^{0.9(1-2\alpha)})}\right) = o\left(n^{-1/2}\right),$$

and the lemma follows. $\square$

Chapter 3

Degree distribution

In this chapter, we will prove Theorem 1.4.1:

Theorem. *Let* $\alpha > \frac{1}{2}$*. Let* $\xi = \frac{4\alpha\nu}{\pi(2\alpha-1)}$.

Let $N_n(k)$ *denote the number of degree* k *vertices in the KPKVB model* $G(n;\alpha,\nu)$ *and consider a sequence of integers* $(k_n)_n$ *with* $0 \leq k_n \leq n-1$.

1. *If* $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$ *as* $n \to \infty$*, then a.a.s.*
$$N_n(k_n) = (1+o(1))np_{k_n},$$
where
$$p_{k_n} = \frac{2\alpha\xi^{2\alpha}\Gamma^+(k_n - 2\alpha, \xi)}{k_n!}.$$

2. *If* $k_n = (1+o(1))cn^{\frac{1}{2\alpha+1}}$ *for some fixed* $c > 0$*, then*
$$N_n(k_n) \xrightarrow[n\to\infty]{d} \mathrm{Po}(2\alpha\xi^{2\alpha}c^{-(2\alpha+1)}).$$

3. *If* $k_n \gg n^{\frac{1}{2\alpha+1}}$*, then a.a.s.* $N_n(k_n) = 0$.

In the proof of statement (i) and (ii) of Theorem 1.4.1, we will eventually work with the Poissonized KPKVB model in the coordinates of the box model, which we denote by $\overline{G_{Po}}$. By this, we mean that the vertex set is given by a Poisson process in the upper half-plane whose intensity function is given by

$$f_n(x,y) = \frac{\alpha\nu}{\pi} e^{-\alpha y} \mathbb{1}_{\{-\frac{\pi n}{2\nu} < x \leq \frac{\pi n}{2\nu}, 0 < y < 2\ln\frac{n}{\nu}\}}$$

and the edges are given by the transformed formula based on the hyperbolic law of cosines, i.e.

$$(x_1,y_1) \sim (x_2,y_2) \Leftrightarrow |x_1 - x_2|_{\frac{\pi n}{\nu}} \leq \Phi(y_1,y_2),$$

where

$$\Phi(y_1,y_2)=\begin{cases}\frac{1}{2}e^{\frac{R}{2}}\arccos\left(\frac{\cosh(R-y_1)\cosh(R-y_2)-\cosh R}{\sinh(R-y_1)\sinh(R-y_2)}\right), & \text{if } y_1+y_2\le R,\\ \frac{\pi}{2}e^{\frac{R}{2}}=\frac{\pi n}{2\nu}, & \text{if } y_1+y_2>R.\end{cases} \tag{3.1}$$

The Poissonization (i.e. the transition from the KPKVB model G_n to the Poissonized KPKVB model G_{Po}) will be justified in Lemma 3.1.10 below and the use of the intensity function f_n (of the box model, i.e. the transition from G_{Po} to $\overline{G_{Po}}$) with the transformed adjacency formula $\Phi(y_1,y_2)$ is justified by the coupling Lemma 1.6.1 and will be done in the proofs of the lemmas whose statements are in terms of G_{Po}.

Let the measure with density f_n be denoted by μ_n.

Denote the intensity measure of the neighbourhood ball of the point $(0,y)$ in $\overline{G_{Po}}$ by

$$\mu_{Po,n}(y)=\int_0^R 2\Phi(y,y_1)\frac{\alpha\nu}{\pi}e^{-\alpha y_1}dy_1. \tag{3.2}$$

Note that we are using the transformed adjacency formula based on the hyperbolic law of cosines. In other words, $\mu_{Po,n}(y)$ is the expected degree of a point with height y in $\overline{G_{Po}}$, the Poissonized KPKVB model in the coordinates of the box model.

Let $\mu(y)=\xi e^{\frac{y}{2}}$.

First of all, if $k_n=k\in\mathbb{N}$ is constant in statement (i) of Theorem 1.4.1, statement (i) has already been shown by Gugelmann et al. [29, Theorem 2.2]. Our main motivation is to extend their result to a complete description of the asymptotic behaviour of $N_n(k_n)$, including all sequences $k_n\to\infty$. In particular, we show that the power-law (with exponent $2\alpha+1$) holds up to the maximum scaling of $k_n=o\left(n^{\frac{1}{2\alpha+1}}\right)$ (up to which we can expect vertices of degree exactly k_n) and derive a Poisson limit distribution in the boundary case $k_n=\Theta(n^{\frac{1}{2\alpha+1}})$. To show this extension, we will assume that $k_n\to\infty$ at several places throughout the proof(s). Although our methods could be adapted to also treat the constant k case, our presentation will assume that $k_n\to\infty$ for the sake of simplicity and clarity.

3.1 Statement of lemmas

We begin by giving an overview of the lemmas needed to prove Theorem 1.4.1. We will group these lemmas thematically such that the last lemma in each subsection is the one which we will use in the main proof of Theorem 1.4.1. The main devision is between small degrees, by which we mean $k_n=O\left(n^{\frac{1}{2\alpha+1}}\right)$, and large degrees, i.e. $k_n\gg n^{\frac{1}{2\alpha+1}}$. For small degrees, we start with the first moment (in the Poissonized KPKVB model), which already yields the limiting expression. Then, we continue with the factorial moments (for the convergence in probability only the second moment is required, but for the Poisson limit law in the boundary case, we do need all factorial moments). We end the section on small degrees with the lemma justifying the Poissonization.

3.1.1 First moment for small degrees

Lemma 3.1.1 (Expected degree given height in poissonized KPKVB)**.** *Let $\alpha > \frac{1}{2}$. Let $\epsilon > 0$ and $0 \le y \le (1-\epsilon)R$. Then as $n \to \infty$, uniformly in y,*

$$\mu_{Po,n}(y) = (1+o(1))\mu(y),$$

and for the derivative w.r.t. y,

$$\mu'_{Po,n}(y) = (1+o(1))\frac{1}{2}\mu(y) = (1+o(1))\frac{1}{2}\mu_{Po,n}(y).$$

Lemma 3.1.2. *Let $0 < \varepsilon < 1$. Then for all $0 \le k_n \le n-1$,*

$$\int_0^{(1-\varepsilon)R} \mathbb{P}\left(\mathrm{Po}(\mu_{Po,n}(y)) = k_n\right) \alpha e^{-\alpha y}\, \mathrm{d}y = (1+o\left(1\right))p_{k_n},$$

where

$$p_{k_n} = \int_0^\infty \mathbb{P}(\mathrm{Po}(\xi e^{\frac{y}{2}}) = k_n)\alpha e^{-\alpha y} dy.$$

Lemma 3.1.3 (First moment of number of degree k vertices)**.** *Let $N_{Po}(k)$ denote the number of degree k vertices in the Poissonized KPKVB model G_{Po} and consider a sequence of integers $k_n \to \infty$ with $0 \le k_n \le n-1$. If $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$, then*

$$\mathbb{E}\left[N_{Po}(k_n)\right] = (1+o(1))np_{k_n}.$$

If $k_n \gg n^{\frac{1}{2\alpha+1}}$, then

$$\mathbb{E}\left[N_{Po}(k_n)\right] = o\left(1\right).$$

3.1.2 Factorial moments for small degrees

For $k \in \mathbb{N}$, $c > 0$, define $y^-_{k,c} = 2\ln\frac{k-c\sqrt{k\ln k}}{\xi}$ (if the argument of the logarithm satisfies $\frac{k-c\sqrt{k\ln k}}{\xi} \le 1$, we set $y^-_{k,c} = 0$), $y^+_{k,c} = 2\ln\frac{k+c\sqrt{k\ln k}}{\xi}$ (recall that as we let $k \to \infty$, we can assume that the argument of the logarithm is ≥ 1) and $\mathcal{S}_{k,c} = \mathcal{R} \cap (\mathbb{R} \times [y^-_{k,c}, y^+_{k,c}])$.

For $r \in \mathbb{N}$, we write $G_{\mathrm{Po}} \cup \{v_1, \ldots, v_r\}$ for the Poissonized KPKVB model obtained by adding $v_1, \ldots, v_r$ to the vertex set of the graph and adding all corresponding edges according to the transformed adjacency formula Φ. Recall that μ_n denotes the intensity measure of the vertex set of G_{Po}.

For a positive integer s, $v_1, \dots, v_s \in \mathcal{R}$ and $V \subset \{v_1, \dots, v_s\}$, define

$$\varphi(V; v_1, \dots, v_s) = \mathbb{P}(\text{every } v \in V \text{ has degree } k_n \text{ in } G_{\mathrm{Po}} \cup \{v_1, \dots, v_s\}).$$

In the following lemmas, let (k_n) be a sequence of integers with $0 \le k_n \le n-1$, $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$ and $k_n \to \infty$.

Lemma 3.1.4 (First moment over strip). *There is $c > 0$ such that*

$$\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1) = (1 + o(1))2\alpha\xi^{2\alpha} n k_n^{-(2\alpha+1)}.$$

Lemma 3.1.5 (Asymptotic factorization of degree probabilities). *Let $C > 0$, $c > 0$. Let r, s be positive integers with $r + 1 \le s$. Fix $0 < \epsilon < 1$. Then, it holds uniformly for all $(v_1, \dots, v_s) \in (\mathcal{S}_{k_n,c})^s$, satisfying $|x_{v_i} - x_{v_{r+1}}| \ge k_n^{1+\epsilon}$ for all $1 \le i \le r$, that*

$$\begin{aligned}
&\varphi(\{v_1, \dots, v_{r+1}\}; v_1, \dots, v_s) \\
&= (1 + o(1))\varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_s)\varphi(\{v_{r+1}\}; v_1, \dots, v_s) + O\left(k_n^{-C}\right).
\end{aligned}$$

(where the uniformity means that there are functions $g_n \in o(1)$ and $h_n \in O\left(k_n^{-C}\right)$, i.e. $g_n(v_1, \dots, v_s) \to 0$ as $n \to \infty$ and $\sup_{n\in\mathbb{N}} h_n(v_1, \dots, v_s) < \infty$, such that the above claim holds simultaneously for all $(v_1, \dots, v_s)$ with the assumed properties as $n \to \infty$, i.e. the speed of the convergence of g_n and the upper bound on h_n do not depend on $v_1, \dots, v_s$.)

Lemma 3.1.6 (Factorization over strips). *There is $c > 0$ such that*

$$\begin{aligned}
&\int_{\mathcal{S}_{k_n,c}} \cdots \int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_r)\mu_n(dv_1) \cdots \mu_n(dv_r) \\
&= (1 + o(1)) \left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)^r.
\end{aligned}$$

Lemma 3.1.7 (Concentration of heights). *Let $r \in \mathbb{N}$. For c sufficiently large,*

$$\begin{aligned}
&\int_{\mathcal{R}} \cdots \int_{\mathcal{R}} \varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_r)\mu_n(dv_1) \cdots \mu_n(dv_r) \\
&= (1 + o(1)) \int_{\mathcal{S}_{k_n,c}} \cdots \int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_r)\mu_n(dv_1) \cdots \mu_n(dv_r).
\end{aligned}$$

Lemma 3.1.8 (Factorization over $\mathcal{R}$). *For every $r \ge 1$,*

$$\begin{aligned}
&\int_{\mathcal{R}} \cdots \int_{\mathcal{R}} \varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_r)\mu_n(dv_1) \cdots \mu_n(dv_r) \\
&= (1 + o(1)) \left(\int_{\mathcal{R}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)^r.
\end{aligned}$$

Lemma 3.1.9 (Factorial moments)**.** *Recall that $N_{Po}(k)$ denotes the number of degree k vertices in G_{Po}. For any positive integer r, it holds that*

$$\mathbb{E}\left[\binom{N_{Po}(k_n)}{r}\right] = (1+o(1))\frac{(\mathbb{E}\left[N_{Po}(k_n)\right])^r}{r!}.$$

3.1.3 Poissonization for small degrees

Let $\alpha > \frac{1}{2}$ and consider a sequence of integers (k_n) with $0 \leq k_n \leq n-1$ and $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$.

Lemma 3.1.10. *As $n \to \infty$, it holds that*

$$\mathbb{E}\left[|N_n(k_n) - N_{Po}(k_n)|\right] = o\left(\mathbb{E}\left[N_{Po}(k_n)\right]\right).$$

3.1.4 First moment for large degrees

Let $\mathrm{Bin}(n,p)$ denote a random variable with a Binomial distribution with n trials and success probability p and denote by $\mathrm{Po}(\lambda)$ a random variable with a Poisson distribution with expectation λ.

Lemma 3.1.11. *Let $n \geq 1$, $0 < \lambda < n$. Then, for any integer $0 \leq k \leq n-1$,*

$$\mathbb{P}\left(\mathrm{Bin}(n,\lambda/n) = k\right) \leq \frac{e}{\sqrt{2\pi}}\sqrt{\frac{n}{n-k}}\mathbb{P}\left(\mathrm{Po}(\lambda) = k\right).$$

Lemma 3.1.12. *Let $\alpha > \frac{1}{2}$. Let $N_n(k)$ denote the number of degree k vertices in the KPKVB model $G(n;\alpha,\nu)$ and consider a sequence of integers (k_n) with $0 \leq k_n \leq n-1$ and $k_n \gg n^{\frac{1}{2\alpha+1}}$. Then, $\mathbb{E}[N_n(k_n)] = o(1)$.*

3.2 Main proof

Proof of Proposition 1.4.1:

(i): First of all, by Lemma 3.1.3,

$$\mathbb{E}\left[N_{Po}(k_n)\right] = (1+o(1))np_{k_n}.$$

By Lemma 3.1.9 with $r=2$, it holds that $\mathbb{E}\left[\binom{N_{Po}(k_n)}{2}\right] = (1+o(1))\frac{(\mathbb{E}[N_{Po}(k_n)])^2}{2}$, which implies

$$\begin{aligned}\mathbb{E}\left[N_{Po}(k_n)^2\right] &= 2\mathbb{E}\left[\binom{N_{Po}(k_n)}{2}\right] + \mathbb{E}\left[N_{Po}(k_n)\right] \\ &= (1+o(1))(\mathbb{E}\left[N_{Po}(k_n)\right])^2 + o((\mathbb{E}\left[N_{Po}(k_n)\right])^2)\end{aligned}$$

(because $\mathbb{E}\left[N_{Po}(k_n)\right] = (1+o(1))np_{k_n} \to \infty$). Hence, by Chebychev for any $\epsilon > 0$,

$$\mathbb{P}(|N_{Po}(k_n) - \mathbb{E}\left[N_{Po}(k_n)\right]| \geq \epsilon\mathbb{E}\left[N_{Po}(k_n)\right]) \leq \frac{Var(N_{Po}(k_n))}{\epsilon^2(\mathbb{E}\left[N_{Po}(k_n)\right])^2} = o(1).$$

As $N_n(k_n) = N_{Po}(k_n)+N_n(k_n)-N_{Po}(k_n) = N_{Po}(k_n)\pm|N_n(k_n)-N_{Po}(k_n)|$ (where the sign depends on whether $N_n(k_n) > N_{Po}(k_n)$ or not), due to Lemma 3.1.10, we have that

$$\begin{aligned}\mathbb{E}\left[N_n(k_n)\right] &= \mathbb{E}\left[N_{Po}(k_n)\right] \pm \mathbb{E}\left[|N_n(k_n) - N_{Po}(k_n)|\right] \\ &= (1+o(1))\mathbb{E}\left[N_{Po}(k_n)\right] = (1+o(1))np_{k_n}.\end{aligned}$$

Finally, note that $p_{k_n} = \frac{2\alpha\xi^{2\alpha}\Gamma^+(k_n-2\alpha,\xi)}{k_n!} = (1+o(1))2\alpha\xi^{2\alpha}k_n^{-(2\alpha+1)}$ as $k_n \to \infty$ as derived in Equations (4.2), (4.3) and (4.4).

(ii): Let $\zeta = 2\alpha\xi^{2\alpha}c^{-(2\alpha+1)} \in \mathbb{R}$. The proof consists of showing that

$$\mathbb{E}\left[\binom{N_{Po}(k_n)}{r}\right] \to \frac{\zeta^r}{r!}$$

for every positive integer r.

If $k_n = (1+o(1))cn^{\frac{1}{2\alpha+1}}$, then by Lemma 3.1.3,

$$\begin{aligned}\mathbb{E}\left[N_{Po}(k_n)\right] &= (1+o(1))2\alpha\xi^{2\alpha}n(1+o(1))^{-(2\alpha+1)}c^{-(2\alpha+1)}n^{-1} \\ &= (1+o(1))2\alpha\xi^{2\alpha}c^{-(2\alpha+1)} = (1+o(1))\zeta,\end{aligned}$$

which implies $\mathbb{E}\left[N_{Po}(k_n)\right] \to \zeta$ (as ζ is a positive constant). From Lemma 3.1.9, it then follows that $\mathbb{E}\left[\binom{N_{Po}(k_n)}{r}\right] = (1+o(1))\frac{(\mathbb{E}[N_{Po}(k_n)])^r}{r!} \to \frac{\zeta^r}{r!}$. Thus, it follows from [6, Theorem 8.3.1] that $N_{Po}(k_n) \xrightarrow{d} Po(\zeta)$ for the Poissonized model. As $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$ and so by Lemma 3.1.10, $\mathbb{E}\left[|N_n(k_n) - N_{Po}(k_n)|\right] = o(\mathbb{E}\left[N_{Po}(k_n)\right]) = o(\zeta)$, it follows that $\mathbb{P}(|N_n(k_n) - N_{Po}(k_n)| \geq 1) \leq \mathbb{E}\left[|N_n(k_n) - N_{Po}(k_n)|\right] = o(\zeta)$. Hence, it also holds that $N_n(k_n) \xrightarrow{d} Po(\zeta)$ in the original KPKVB model.

(iii): In this case we have by Lemma 3.1.12 that $\mathbb{E}[N_n(k_n)] = o(1)$. Hence by Markov's inequality (resp. the first moment method),

$$\mathbb{P}(N_n(k_n) > 0) \leq \mathbb{E}[N_n(k_n)] = o(1).$$

■

3.3 Proofs of the lemmas

3.3.1 Implications of Chernoff bound

We recall the following version of the Chernoff bound (a proof can for instance be found in [44]):

Lemma 3.3.1 (Chernoff bound)**.** *For a Poisson random variable X with expectation μ, for $k \in \mathbb{N}$, if $\mu > k$, then*

$$\mathbb{P}(X \leq k) \leq e^{-\mu H\left(\frac{k}{\mu}\right)} \tag{3.3}$$

and if $\mu < k$, then

$$\mathbb{P}(X \geq k) \leq e^{-\mu H\left(\frac{k}{\mu}\right)} \tag{3.4}$$

where $H(x) = x \ln x - x + 1$.

Lemma 3.3.2 (Implications of Chernoff bound)**.** *Let $C > 0$. Then, there is $c > 0$ such that for all $y \notin [y^-_{k,c}, y^+_{k,c}]$ (for the definition of $y^-_{k,c}$ and $y^+_{k,c}$ see the first sentence of Subsection 3.1.2), we have that as $k \to \infty$,*

$$\mathbb{P}(Po(\mu(y)) = k) = O\left(k^{-C}\right).$$

Proof. If $y > y^+_{k,c}$, then $\mu(y) > k + c\sqrt{k \ln k} > k$. The event $\{Po(\mu(y)) = k\}$ implies the event $\{Po(\mu(y)) \leq k\}$, so

$$\mathbb{P}(Po(\mu(y)) = k) \leq \mathbb{P}(Po(\mu(y)) \leq k).$$

As $\mu(y) > k+c\sqrt{k \ln k}$, it holds that $\frac{k}{\mu(y)} < \frac{k}{k+c\sqrt{k \ln k}} < 1$. As $H(x) = x\ln x - x + 1$ is decreasing for $x < 1$, it follows that $H\left(\frac{k}{\mu(y)}\right) > H\left(\frac{k}{k+c\sqrt{k\ln k}}\right)$. By the Chernoff bound, see (3.3) in Lemma 3.3.1,

$$\begin{aligned}\mathbb{P}(Po(\mu(y)) \leq k) &\leq e^{-\mu(y) H\left(\frac{k}{\mu(y)}\right)} \\ &\leq e^{-(k+c\sqrt{k\ln k}) H\left(\frac{k}{k+c\sqrt{k\ln k}}\right)}.\end{aligned}$$

Now, we use that for $x < 1$, $H(x) \geq \frac{1}{2}(x-1)^2 x$, to obtain,

$$\mathbb{P}(Po(\mu(y)) \leq k) \leq e^{-\frac{1}{2}(k+c\sqrt{k\ln k})\left(\frac{k}{k+c\sqrt{k\ln k}}-1\right)^2 \frac{k}{k+c\sqrt{k\ln k}}} = e^{-\frac{c^2k^2\ln k}{2(k+c\sqrt{k\ln k})^2}}.$$

Next, $k + c\sqrt{k\ln k} \leq 2k$ (for k large enough) implies that by picking $c = \sqrt{8C}$,

$$\mathbb{P}(Po(\mu(y)) \leq k) \leq e^{-\frac{1}{2}\frac{c^2 \ln k}{4}} = k^{-C}.$$

If $y < y^-_{k,c}$, then $\mu(y) < k - c\sqrt{k\ln k} < k$. The event $\{Po(\mu(y)) = k\}$ implies the event $\{Po(\mu(y)) \geq k\}$, so

$$\mathbb{P}(Po(\mu(y)) = k) \leq \mathbb{P}(Po(\mu(y)) \geq k).$$

As $\mu(y) < k-c\sqrt{k\ln k}$, it holds that $\frac{k}{\mu(y)} > \frac{k}{k-c\sqrt{k\ln k}} > 1$ (note that $k-c\sqrt{k\ln k} > 0$ for k large enough). As $H(x) = x\ln x - x + 1$ is increasing for $x > 1$, it follows that $H\left(\frac{k}{\mu(y)}\right) > H\left(\frac{k}{k-c\sqrt{k\ln k}}\right)$. By the Chernoff bound, see (3.4) in Lemma 3.3.1,

$$\mathbb{P}(Po(\mu(y)) \geq k) \leq e^{-\mu(y) H\left(\frac{k}{\mu(y)}\right)}.$$

Now, we use that for $x > 1$, $H(x) \geq \frac{1}{4}(x-1)^2$, to obtain,

$$\mathbb{P}(Po(\mu(y)) \geq k) \leq e^{-\frac{1}{4}\mu(y)\left(\frac{k}{\mu(y)}-1\right)^2} = e^{-\frac{1}{4}\frac{(k-\mu(y))^2}{\mu(y)}}.$$

As $\mu(y) < k - c\sqrt{k \ln k}$, it follows that $k - \mu(y) > c\sqrt{k \ln k} > 0$ and hence, $(k-\mu(y))^2 > c^2 k \ln k$. Together with $\frac{k}{\mu(y)} > 1$, this yields

$$\mathbb{P}(Po(\mu(y)) \geq k) \leq e^{-\frac{c^2 k \ln k}{4\mu(y)}} \leq e^{-\frac{c^2}{4}\ln k} \leq e^{-\frac{c^2}{8}\ln k}.$$

Finally, we note that this is $\leq k^{-C}$ by our choice of $c = \sqrt{8C}$. □

3.3.2 Expected degree given height in Poissonized KPKVB

We provide two different proofs for the first part of Lemma 3.1.1.

Proof of Lemma 3.1.1, first part: Recall from Lemma 1.6.2 that there is a positive constant K such that for any $\epsilon > 0$, for all $y_1, y_2 \in [0, (1-\epsilon)R]$, $y_1 + y_2 < R$, we have (for $\Phi(y_1, y_2)$ from (3.1)) the estimates,

$$e^{\frac{1}{2}(y_1+y_2)} - Ke^{\frac{3}{2}(y_1+y_2)-R} \leq \Phi(y_1, y_2) \leq e^{\frac{1}{2}(y_1+y_2)} + Ke^{\frac{3}{2}(y_1+y_2)-R}.$$

Recall the definition of $\mu_{Po,n}(y)$ from (3.2). Due to the case distinction in the definition of $\Phi(y, y_1)$, we split the integral for $\mu_{Po,n}(y)$ into two integrals I_1 and I_2, i.e.

$$\begin{aligned}\mu_{Po,n}(y) &= \int_0^R 2\Phi(y, y_1)\frac{\alpha\nu}{\pi}e^{-\alpha y_1}dy_1 \\ &= \int_0^{R-y} 2\Phi(y, y_1)\frac{\alpha\nu}{\pi}e^{-\alpha y_1}dy_1 + \int_{R-y}^R \frac{\pi n}{\nu}\frac{\alpha\nu}{\pi}e^{-\alpha y_1}dy_1 =: I_1 + I_2.\end{aligned}$$

Firstly, we will show that the second integral $I_2 = o(\mu(y))$ and then we will show that $I_1 = (1+o(1))\mu(y)$ (both with convergence uniform in $0 \leq y \leq (1-\epsilon)R$).

For the second integral I_2, we compute

$$\begin{aligned}I_2 &= \int_{R-y}^R \frac{\pi n}{\nu}\frac{\alpha\nu}{\pi}e^{-\alpha y_1}dy_1 \\ &= n(e^{-\alpha(R-y)} - e^{-\alpha R}) = ne^{-\alpha R}(e^{\alpha y} - 1) = n^{1-2\alpha}\nu^{2\alpha}(e^{\alpha y} - 1).\end{aligned}$$

To see that $n^{1-2\alpha}\nu^{2\alpha}(e^{\alpha y} - 1) = o(\mu(y))$, recall that $\mu(y) = \xi e^{\frac{y}{2}}$. So, we need to show that

$$\frac{n^{1-2\alpha}\nu^{2\alpha}(e^{\alpha y} - 1)}{\xi e^{\frac{y}{2}}} = o(1)$$

or equivalently that

$$\frac{e^{\alpha y}-1}{\xi e^{\frac{y}{2}}} = o\left(n^{2\alpha-1}\right).$$

For this, note that

$$\frac{e^{\alpha y}-1}{\xi e^{\frac{y}{2}}} = O\left(e^{\alpha y-\frac{y}{2}}\right) = O\left(e^{\left(\alpha-\frac{1}{2}\right)y}\right).$$

As $y \leq (1-\epsilon)R = (1-\epsilon)2\ln\frac{n}{\nu}$ and $\alpha > \frac{1}{2}$, we have

$$e^{\left(\alpha-\frac{1}{2}\right)y} \leq e^{\left(\alpha-\frac{1}{2}\right)(1-\epsilon)R} = \left(\frac{n}{\nu}\right)^{2\left(\alpha-\frac{1}{2}\right)(1-\epsilon)} = o\left(n^{2\alpha-1}\right),$$

where the convergence is uniform in y, $0 \leq y \leq (1-\epsilon)R$, as the last upper bound does not depend on y.

For the first integral I_1, define a main and error term by

$$\begin{aligned}
I_{1,main} &= \int_0^{R-y} 2e^{\frac{y+y_1}{2}} \frac{\alpha\nu}{\pi} e^{-\alpha y_1} dy_1, \\
I_{1,error} &= \int_0^{R-y} 2Ke^{\frac{3}{2}(y+y_1)-R} \frac{\alpha\nu}{\pi} e^{-\alpha y_1} dy_1.
\end{aligned}$$

From the error bounds for Φ as given in Lemma 1.6.2, it follows that

$$I_{1,main} - I_{1,error} \leq I_1 \leq I_{1,main} + I_{1,error}.$$

We will firstly show that $I_{1,main} = (1+o(1))\mu(y)$ and then that $I_{1,error} = o(\mu(y))$.

For the main term, we obtain, as $R - y \geq \epsilon R \to \infty$, uniformly in $0 \leq y \leq (1-\epsilon)R$,

$$\begin{aligned}
I_{1,main} = \int_0^{R-y} 2e^{\frac{y+y_1}{2}} \frac{\alpha\nu}{\pi} e^{-\alpha y_1} dy_1 &= \frac{2\alpha\nu}{\pi} e^{\frac{y}{2}} \int_0^{R-y} e^{\left(\frac{1}{2}-\alpha\right)y_1} dy_1 \\
&= \frac{2\alpha\nu}{\pi\left(\alpha-\frac{1}{2}\right)} e^{\frac{y}{2}} \left(1 - e^{\left(\frac{1}{2}-\alpha\right)(R-y)}\right) \\
&= (1+o(1))\xi e^{\frac{y}{2}} = (1+o(1))\mu(y).
\end{aligned}$$

For the error term, we obtain, for $\alpha \neq \frac{3}{2}$, uniformly in $0 \leq y \leq (1-\epsilon)R$,

$$\begin{aligned}
I_{1,error} &= \int_0^{R-y} 2Ke^{\frac{3}{2}(y+y_1)-R} \frac{\alpha\nu}{\pi} e^{-\alpha y_1} dy_1 = 2K\frac{\alpha\nu}{\pi} e^{\frac{3}{2}y-R} \int_0^{R-y} e^{\left(\frac{3}{2}-\alpha\right)y_1} dy_1 \\
&= \frac{2K\alpha\nu}{\pi\left(\frac{3}{2}-\alpha\right)} e^{\frac{3}{2}y-R} \left(e^{\left(\frac{3}{2}-\alpha\right)(R-y)} - 1\right)
\end{aligned}$$

$$= \frac{2K\alpha\nu}{\pi\left(\frac{3}{2}-\alpha\right)} e^{\frac{1}{2}y}\left(e^{\left(\frac{1}{2}-\alpha\right)(R-y)} - e^{-(R-y)}\right) = o\left(\xi e^{\frac{y}{2}}\right).$$

For the error term with $\alpha = \frac{3}{2}$, uniformly in $0 \le y \le (1-\epsilon)R$,

$$\begin{aligned}\int_0^{R-y} 3Ke^{\frac{3}{2}(y+y_1)-R}\frac{\nu}{\pi}e^{-\frac{3}{2}y_1}\,dy_1 &= 3K\frac{\nu}{\pi}e^{\frac{3}{2}y-R}\int_0^{R-y} dy_1 \\ &= \frac{3K\nu}{\pi}e^{\frac{3}{2}y-R}(R-y) = o\left(\xi e^{\frac{y}{2}}\right).\end{aligned}$$

We conclude that $I_{1,error} = o(\mu(y))$ and hence $I_{1,main} \pm I_{1,error} = (1+o(1))\mu(y)$, which finishes the proof. ■

In order to prove the second part of Lemma 3.1.1 and for later use, we will show the following auxiliary lemma:

Lemma 3.3.3. *Let $\Phi(y,y')$ be defined as in (1.8) and $\mathcal{B}_\infty(y)$ the ball around p in the infinite limit model G_∞ as in (1.4). Then, for any $0 \le \delta < 1$,*

$$\lim_{n\to\infty}\sup_{0\le y\le(1-\varepsilon)R}\left|\mu\left(\mathcal{B}_\infty(y)\right)^{-1}\frac{2\nu\alpha}{\pi}\int_0^{(1-\delta)(R-y)}\Phi(y,y')e^{-\alpha y'}\,\mathrm{d}y' - 1\right| = 0.$$

Proof. Recall that $\mu\left(\mathcal{B}_\infty(0,y)\right) = \xi e^{y/2}$ where $\xi = \frac{4\alpha\nu}{\pi(2\alpha-1)}$ and, by Lemma 1.6.2,

$$\left|\Phi(y,y') - e^{\frac{y+y'}{2}}\right| \le Ke^{\frac{3}{2}(y+y')-R},$$

for all $y+y' < R$. Also,

$$\mu\left(\mathcal{B}_\infty(y)\right) = \frac{2\nu\alpha}{\pi}\int_0^\infty e^{\frac{y+y'}{2}}e^{-\alpha y'}\,\mathrm{d}y'.$$

Therefore we have

$$\begin{aligned}&\left|\mu\left(\mathcal{B}_\infty(y)\right)^{-1}\frac{2\nu\alpha}{\pi}\int_0^{(1-\delta)(R-y)}\Phi(y,y')e^{-\alpha y'}\,\mathrm{d}y' - 1\right| \\ &= \frac{1}{\mu\left(\mathcal{B}_\infty(y)\right)}\left|\frac{2\nu\alpha}{\pi}\int_0^{(1-\delta)(R-y)}\Phi(y,y')e^{-\alpha y'}\,\mathrm{d}y' - \frac{2\nu\alpha}{\pi}\int_0^\infty e^{\frac{y+y'}{2}}e^{-\alpha y'}\,\mathrm{d}y'\right| \\ &\le \frac{2\nu\alpha}{\pi\mu\left(\mathcal{B}_\infty(y)\right)}\int_0^{(1-\delta)(R-y)}\left|\Phi(y,y') - e^{\frac{y+y'}{2}}\right|e^{-\alpha y'}\,\mathrm{d}y' \\ &\qquad + \frac{2\nu\alpha}{\pi\mu\left(\mathcal{B}_\infty(y)\right)}\int_{(1-\delta)(R-y)}^\infty e^{\frac{y+y'}{2}}e^{-\alpha y'}\,\mathrm{d}y' \\ &\le \frac{2K\nu\alpha}{\pi\mu\left(\mathcal{B}_\infty(y)\right)}e^{\frac{3}{2}y-R}\int_0^{(1-\delta)(R-y)}e^{\left(\frac{3}{2}-\alpha\right)y'}\,\mathrm{d}y' + e^{-\left(\alpha-\frac{1}{2}\right)(1-\delta)(R-y)}.\end{aligned}$$

For the last term it holds that

$$\lim_{n\to\infty} \sup_{0<y\leq(1-\varepsilon)R} e^{-(\alpha-\frac{1}{2})(1-\delta)(R-y)} = 0.$$

We first compute the integral in the first term, which depends on the value of α,

$$\int_0^{(1-\delta)(R-y)} e^{(\frac{3}{2}-\alpha)y'} \,\mathrm{d}y' = \begin{cases} \frac{2}{3-2\alpha}\left(e^{(\frac{3}{2}-\alpha)(1-\delta)(R-y)} - 1\right) & \text{if } 1/2<\alpha<3/2, \\ (1-\delta)(R-y) & \text{if } \alpha = 3/2, \\ \frac{2}{2\alpha-3}\left(1 - e^{-(\alpha-\frac{3}{2})(1-\delta)(R-y)}\right) & \text{if } \alpha > 3/2. \end{cases}$$

This implies that

$$\begin{aligned} &\frac{2K\nu\alpha}{\pi\mu\left(\mathcal{B}_\infty\left(y\right)\right)} e^{\frac{3y}{2}-R} \int_0^{(1-\delta)(R-y)} e^{(\frac{3}{2}-\alpha)y'} \,\mathrm{d}y' \\ &= \begin{cases} \frac{(2\alpha-1)K}{3-2\alpha}\left(e^{-(\alpha-\frac{1}{2})(R-y)-(\frac{3}{2}-\alpha)\delta(R-y)} - e^{-(R-y)}\right) & \text{if } 1/2<\alpha<3/2, \\ \frac{(2\alpha-1)K}{2}(1-\delta)(R-y)e^{-(R-y)} & \text{if } \alpha = 3/2, \\ \frac{(2\alpha-1)K}{2\alpha-3}\left(e^{-(R-y)} - e^{-(\alpha-\frac{1}{2})(R-y)-(\alpha-\frac{3}{2})(R-y)}\right) & \text{if } \alpha > 3/2, \end{cases} \end{aligned}$$

and hence

$$\lim_{n\to\infty} \sup_{0\leq y\leq(1-\varepsilon)R} \frac{2K\nu\alpha}{\pi\mu\left(\mathcal{B}_\infty\left(y\right)\right)} e^{\frac{3y}{2}-R} \int_0^{(1-\delta)(R-y)} e^{(\frac{3}{2}-\alpha)y'} \,\mathrm{d}y' = 0,$$

which completes the proof. □

The following lemma provides the second proof of the first part of Lemma 3.1.1:

Lemma 3.3.4. *For any* $0<\varepsilon<1$,

$$\lim_{n\to\infty} \sup_{0\leq y\leq(1-\varepsilon)R} \left| \frac{\mu\left(\mathcal{B}\left(y\right)\right)}{\mu\left(\mathcal{B}_\infty\left(y\right)\right)} - 1 \right| = 0.$$

Proof. We perform the computation of $\mu\left(\mathcal{B}\left(y\right)\right)$ by splitting the integration with respect to the height y' into the cases $y' > R-y$ and $y' \leq R-y$, i.e.

$$\mu\left(\mathcal{B}\left(y\right)\right) = \mu\left(\mathcal{B}\left(y\right)\cap\mathcal{R}([0,R-y))\right) + \mu\left(\mathcal{B}\left(y\right)\cap\mathcal{R}([R-y,R])\right).$$

For the first part we have that

$$\mu\left(\mathcal{B}\left((0,y)\right)\cap\mathcal{R}[(0,R-y)]\right) = \frac{2\nu\alpha}{\pi}\int_0^{R-y} \Phi(y,y')e^{-\alpha y'} \,\mathrm{d}y'.$$

Hence, by applying Lemma 3.3.3 with $\delta = 0$, we conclude that

$$\lim_{n\to\infty} \sup_{0\leq y\leq(1-\varepsilon)R} \left| \frac{\mu\left(\mathcal{B}\left((y)\right)\cap\mathcal{R}[(0,R-y)]\right)}{\mu\left(\mathcal{B}_\infty\left(y\right)\right)} - 1 \right| = 0.$$

For the second part we observe that $\mathcal{B}\left((y)\right) \cap \mathcal{R}([R-y,R]) = \mathcal{R}([R-y,R])$. Thus, recalling that $f_{\alpha,\nu}$ is the intensity function of the infinite limit model as defined in (1.3),

$$\begin{aligned}
&\mu\left(\mathcal{B}\left((y)\right) \cap \mathcal{R}([R-y,R])\right) \\
&\qquad = \int_{R-y}^{R} \int_{-\frac{\pi}{2}e^{\frac{R}{2}}}^{\frac{\pi}{2}e^{\frac{R}{2}}} f_{\alpha,\nu}(x',y') \,\mathrm{d}x' \,\mathrm{d}y' = \nu\alpha e^{R/2}\left(e^{-\alpha(R-y)} - e^{-\alpha R}\right) \\
&\qquad = \mu\left(\mathcal{B}_{\infty}\left(y\right)\right) \frac{2\alpha-1}{4\pi}\left(e^{-(\alpha-\frac{1}{2})(R-y)} - e^{-(\alpha-\frac{1}{2})R-y/2}\right),
\end{aligned} \tag{3.5}$$

from which we conclude that

$$\lim_{n\to\infty} \sup_{0\le y\le(1-\varepsilon)R} \frac{\mu\left(\mathcal{B}\left((0,y)\right) \cap \mathcal{R}([R-y,R])\right)}{\mu\left(\mathcal{B}_{\infty}\left(y\right)\right)} = 0,$$

which finishes the proof. $\square$

The following lemma will provide the proof of the second part of Lemma 3.1.1:

Lemma 3.3.5. *For any* $0 < \varepsilon < 1$,

$$\lim_{n\to\infty} \sup_{0\le y\le(1-\varepsilon)R} \left| \mu\left(\mathcal{B}_{\infty}\left(y\right)\right)^{-1} \frac{\partial}{\partial y}\mu\left(\mathcal{B}\left(y\right)\right) - \frac{1}{2}\right| = 0.$$

Proof. We again split $\mu\left(\mathcal{B}\left(y\right)\right)$ over the top and bottom part, as

$$\mu\left(\mathcal{B}\left(y\right)\right) = \mu\left(\mathcal{B}\left(y\right) \cap \mathcal{R}([0,R-y))\right) + \mu\left(\mathcal{B}\left(y\right) \cap \mathcal{R}([R-y,R])\right),$$

where

$$\mu\left(\mathcal{B}\left(y\right) \cap \mathcal{R}([0,R-y))\right) = \frac{2\alpha\nu}{\pi}\int_0^{R-y} \Phi(y,y')e^{-\alpha y'}\,\mathrm{d}y',$$

with $\Phi(y,y')$ defined as in (1.8) and, see (3.5),

$$\mu\left(\mathcal{B}\left(y\right) \cap \mathcal{R}([R-y,R])\right) = \xi e^{y/2}\frac{2\alpha-1}{4\pi}\left(e^{-(\alpha-\frac{1}{2})(R-y)} - e^{-(\alpha-\frac{1}{2})R-y/2}\right).$$

Taking the derivative of the last expression with respect to y gives

$$\begin{aligned}
&\frac{\partial}{\partial y}\mu\left(\mathcal{B}\left(y\right) \cap \mathcal{R}([R-y,R])\right) \\
&= \frac{1}{2}\mu\left(\mathcal{B}\left(y\right) \cap \mathcal{R}([R-y,R])\right) \\
&\qquad\qquad + \xi e^{y/2}\frac{2\alpha-1}{4\pi}\left(\left(\alpha-\frac{1}{2}\right)e^{-(\alpha-\frac{1}{2})(R-y)} + \frac{1}{2}e^{-(\alpha-\frac{1}{2})R-y/2}\right) \\
&= \frac{1}{2}\mu\left(\mathcal{B}\left(y\right) \cap \mathcal{R}([R-y,R])\right)\left(1 + \frac{(2\alpha-1)e^{-(\alpha-\frac{1}{2})(R-y)} + e^{-(\alpha-\frac{1}{2})R-y/2}}{e^{-(\alpha-\frac{1}{2})(R-y)} - e^{-(\alpha-\frac{1}{2})R-y/2}}\right).
\end{aligned}$$

Since, $\lim_{n\to\infty}\sup_{0<y\le(1-\varepsilon)R}\mu\left(\mathcal{B}_\infty(y)\right)^{-1}\mu\left(\mathcal{B}(y)\cap\mathcal{R}([R-y,R])\right)=0$, we are left to show that

$$\lim_{n\to\infty}\sup_{0<y\le(1-\varepsilon)R}\left|\mu\left(\mathcal{B}_\infty(y)\right)^{-1}\frac{2\alpha\nu}{\pi}\frac{\partial}{\partial y}\int_0^{R-y}\Phi(y,y')e^{-\alpha y'}\,\mathrm{d}y'-\frac{1}{2}\right|=0. \quad (3.6)$$

We start with some preliminary computations. For convenience we define

$$\Xi(y,y')=1-\frac{\cosh(R-y)\cosh(R-y')-\cosh(R)}{\sinh(R-y)\sinh(R-y')},$$

so that

$$\Phi(y,y')=\frac{1}{2}e^{R/2}\arccos\left(1-\Xi(y,y')\right).$$

Next, following the same calculation as in the proof of [25, Lemma 28], we write

$$\begin{aligned}\Xi(y,y')&=2e^{-(R-y-y')}\frac{\left(1-e^{y'-y-R}\right)\left(1-e^{y-y'-R}\right)}{\left(1-e^{-2(R-y')}\right)\left(1-e^{-2(R-y)}\right)}\\&:=2e^{-(R-y-y')}\frac{h_1(y)h_2(y)}{h_3(y')h_3(y)},\end{aligned}$$

with

$$h_1(y)=1-e^{y'-y-R},\quad h_2(y)=1-e^{y-y'-R}\quad\text{and}\quad h_3(y)=1-e^{-2(R-y)}.$$

We suppressed the dependence on n and, in some cases, on y' for notation convenience.

We make two important observations. First, $\Xi(y,y')$ is an increasing function in both arguments, for $y,y'<R$ and $y+y'<R$. Second, for all $y+y'<R$, $h_1(y)\le h_3(y')$ and $h_2(y)\le h_3(y)$, while $h_3(y),h_3(y')<1$, so that

$$2e^{-(R-y-y')}h_1(y)h_2(y)\le\Xi(y,y')\le 2e^{-(R-y-y')}. \quad (3.7)$$

In particular, since $R-y$ is an increasing function of n uniformly for $0<y<(1-\varepsilon)R$, there exists a $0<\delta<1$ such that $1/2\le\Xi(y,y')<2$ for all $y+y'<R$ and $(1-\delta)(R-y)<y'<R$ and n large enough.

Next, taking the derivative of $\Xi(y,y')$ yields

$$\begin{aligned}&\frac{\partial}{\partial y}\Xi(y,y')\\&\quad=\Xi(y,y')+2e^{-(R-y-y')}\left(\frac{h_1'(y)h_2(y)}{h_3(y')h_3(y)}+\frac{h_1(y)h_2'(y)}{h_3(y')h_3(y)}-\frac{h_1(y)h_2(y)h_3'(y)}{h_3(y')h_3(y)^2}\right)\\&\quad=\Xi(y,y')\left(1+\frac{h_1'(y)}{h_1(y)}+\frac{h_2'(y)}{h_2(y)}-\frac{h_3'(y)}{h_3(y)}\right)\end{aligned}$$

$$:= \Xi(y,y')\left(1+\varphi_n(y,y')\right),$$

with

$$\varphi_n(y,y') = \frac{e^{y'-y-R}}{1-e^{y'-y-R}} - \frac{e^{y-y'-R}}{1-e^{y-y'-R}} - \frac{2e^{-2(R-y)}}{1-e^{-2(R-y)}}.$$

Therefore, by the chain rule,

$$\begin{aligned}\frac{\partial}{\partial y}\Phi(y,y') &= \frac{1}{2}e^{R/2}\frac{1}{\sqrt{1-(1-\Xi(y,y'))^2}}\frac{\partial}{\partial y}\Xi(y,y') \\ &= \frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-(1-\Xi(y,y'))^2}}\left(1+\varphi_n(y,y')\right). \qquad (3.8)\end{aligned}$$

Applying the Leibniz's rule we then get

$$\begin{aligned}&\frac{\partial}{\partial y}\int_0^{R-y}\Phi(y,y')e^{-\alpha y'}\,\mathrm{d}y' \\ &= -\Phi(y,R-y)e^{-\alpha(R-y)} + \int_0^{R-y}\frac{\partial}{\partial y}\Phi(y,y')e^{-\alpha y'}\,\mathrm{d}y' \\ &= -\frac{1}{2}e^{-(\alpha-\frac{1}{2})R+\alpha y} + \int_0^{R-y}\frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-(1-\Xi(y,y'))^2}}\left(1+\varphi_n(y,y')\right)e^{-\alpha y'}\,\mathrm{d}y' \\ &= -\frac{1}{2}e^{-(\alpha-\frac{1}{2})R+\alpha y} + \int_0^{(1-\delta)(R-y)}\frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-(1-\Xi(y,y'))^2}}\left(1+\varphi_n(y,y')\right)e^{-\alpha y'}\,\mathrm{d}y' \\ &\quad + \int_{(1-\delta)(R-y)}^{R-y}\frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-(1-\Xi(y,y'))^2}}\left(1+\varphi_n(y,y')\right)e^{-\alpha y'}\,\mathrm{d}y' \\ &=: -I_1(y) + I_2(y) + I_3(y),\end{aligned}$$

with $0 \le \delta < 1$ such that $0 < \Xi(y,y') < 2$ for all $0 < y < R$ and $(1-\delta)(R-y) < y' < R$.

We proceed by showing that

$$\lim_{n\to\infty}\sup_{0<y\le(1-\varepsilon)R}\left|\frac{I_t(y)}{\mu\left(\mathcal{B}_\infty\left(y\right)\right)}\right| = 0, \quad \text{for } t=1,3, \qquad (3.9)$$

while

$$\lim_{n\to\infty}\sup_{0\le y\le(1-\varepsilon)R}\left|\frac{2\nu\alpha}{\pi\mu\left(\mathcal{B}_\infty\left(y\right)\right)}I_2(y) - \frac{1}{2}\right| = 0. \qquad (3.10)$$

This then implies (3.6) and finishes the proof.

For $I_1(y)$ we have

$$\lim_{n\to\infty}\sup_{0<y\le(1-\varepsilon)R}\mu\left(\mathcal{B}_\infty\left(y\right)\right)^{-1}I_1(y) \le \lim_{n\to\infty}\sup_{0<y\le(1-\varepsilon)R}\frac{1}{2\xi}e^{-(\alpha-\frac{1}{2})(R-y)} = 0.$$

For $I_3(y)$ we first use that $y' < R - y$ to bound $\varphi(y, y')$ as

$$\varphi_n(y, y') \leq \frac{e^{y'-y-R}}{1 - e^{y'-y-R}} \leq \frac{e^{-2y}}{1 - e^{-2y}}.$$

This then yields that

$$I_3(y) \leq \frac{1}{2}\left(1 + \frac{e^{-2y}}{1 - e^{-2y}}\right) e^{R/2} \int_{(1-\delta)(R-y)}^{R-y} \frac{\Xi(y, y')}{\sqrt{1 - (1 - \Xi(y, y'))^2}} e^{-\alpha y'} \, \mathrm{d}y'.$$

To bound the integral we recall that $0 < \Xi(y, y') \leq 2e^{-(R-y-y')} < 2$ and for all $1/2 \leq x < 2$,

$$\frac{1}{\sqrt{1 - (1 - x)^2}} \leq \frac{2}{\sqrt{2 - x}},$$

where the right-hand side is a monotone increasing function. Therefore

$$\begin{aligned}
&\int_{(1-\delta)(R-y)}^{R-y} \frac{\Xi(y, y')}{\sqrt{1 - (1 - \Xi(y, y'))^2}} e^{-\alpha y'} \, \mathrm{d}y' \\
&\leq 2 \int_{(1-\delta)(R-y)}^{R-y} \frac{\Xi(y, y')}{\sqrt{(2 - \Xi(y, y'))}} e^{-\alpha y'} \, \mathrm{d}y' \\
&\leq \sqrt{2} e^{-\alpha(R-y)} \int_{(1-\delta)(R-y)}^{R-y} \frac{e^{-(R-y-y')}}{\sqrt{1 - e^{-(R-y-y')}}} e^{\alpha(R-y-y')} \, \mathrm{d}y'.
\end{aligned}$$

Making the change of variables $z = e^{-(R-y-y')}$ ($\mathrm{d}y' = z^{-1} \, \mathrm{d}z$) we get that

$$\begin{aligned}
&\sqrt{2} e^{-\alpha(R-y)} \int_{(1-\delta)(R-y)}^{R-y} \frac{e^{-(R-y-y')}}{\sqrt{1 - e^{-(R-y-y')}}} e^{\alpha(R-y-y')} \, \mathrm{d}y' \\
&= \sqrt{2} e^{-\alpha(R-y)} \int_{e^{-\delta(R-y)}}^{1} \frac{z^{-\alpha}}{\sqrt{1 - z}} \, \mathrm{d}z \leq \sqrt{2} e^{-\alpha(R-y)} \sqrt{1 - e^{-\delta(R-y)}} \leq \sqrt{2} e^{-\alpha(R-y)}.
\end{aligned}$$

We therefore conclude that

$$I_3(y) \leq \frac{1}{\sqrt{2}}\left(1 + \frac{e^{-2y}}{1 - e^{-2y}}\right) e^{-(\alpha - \frac{1}{2})R + \alpha y},$$

which implies (3.9) for $t = 3$.

Finally, to show (3.10) we first write

$$\begin{aligned}
\left|\frac{2\alpha\nu}{\pi\mu\left(\mathcal{B}_\infty\left(y\right)\right)} I_2(y) - \frac{1}{2}\right| &\leq \left|\frac{2\alpha\nu}{\pi\mu\left(\mathcal{B}_\infty\left(y\right)\right)} \int_0^{(1-\delta)(R-y)} \frac{\Phi(y, y')}{2} e^{-\alpha y'} \, \mathrm{d}y' - \frac{1}{2}\right| \\
&\quad + \frac{2\alpha\nu}{\pi\mu\left(\mathcal{B}_\infty\left(y\right)\right)} \left|\int_0^{(1-\delta)(R-y)} \frac{\Phi(y, y')}{2} e^{-\alpha y'} \, \mathrm{d}y' - I_2(y)\right|.
\end{aligned}$$

By Lemma 3.3.3,

$$\lim_{n\to\infty} \sup_{0<y\le(1-\varepsilon)R} \left| \frac{2\nu\alpha}{\pi\mu\left(\mathcal{B}_\infty\left(y\right)\right)} \int_0^{(1-\delta)(R-y)} \Phi(y,y')\alpha e^{-\alpha y'}\, \mathrm{d}y' - 1 \right| = 0,$$

and thus it suffices to show that the $\lim_{n\to\infty}\sup_{0<y\le(1-\varepsilon)R}$ of the second term goes to zero.

Recalling the definition of $I_2(y)$ we have

$$\begin{aligned}
&\left| \int_0^{(1-\delta)(R-y)} \frac{\Phi(y,y')}{2} e^{-\alpha y'}\, \mathrm{d}y' - I_2(y) \right| \\
&\le \int_0^{(1-\delta)(R-y)} \left| \frac{\Phi(y,y')}{2} - \frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-\left(1-\Xi(y,y')\right)^2}}(1+\varphi_n(y,y') \right| e^{-\alpha y'}\, \mathrm{d}y'. \quad (3.11)
\end{aligned}$$

We will proceed to bound the term inside the integral. For this we first note that for $0 \le y' \le (1-\delta)(R-y)$,

$$\varphi_n(y,y') \le \frac{e^{-\delta(R-y)}}{1-e^{-\delta(R-y)}},$$

and recall that $\Xi(y,y') \le 2e^{-(R-y-y')}$. Moreover, since $x/\sqrt{1-(1-x)^2} = x/\sqrt{2x-x^2}$ is an increasing function and $e^{-(R-y-y')} \le e^{-\delta(R-y)}$ for $0 < y' < (1-\delta)(R-y)$,

$$\frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-\left(1-\Xi(y,y')\right)^2}} \le e^{R/2}\frac{e^{-\delta(R-y)}}{\sqrt{2e^{-\delta(R-y)}-e^{-2\delta(R-y)}}}.$$

Next, recall that $\Phi(y,y') = \frac{1}{2}e^{R/2}\arccos(1-\Xi(y,y'))$. Then, since $\Xi(y,y') < 1$ for all $y' < (1-\delta)(R-y)$, $y < R$ and n large enough, we have (see Lemma C.1),

$$\left| \frac{1}{2}\Phi(y,y') - \frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-\left(1-\Xi(y,y')\right)^2}} \right| \le \frac{1}{2}\Phi(y,y')\Xi(y,y'),$$

for all $y' < (1-\delta)(R-y)$ and $y < R$. Together these facts imply that for n large enough

$$\begin{aligned}
&\left| \frac{\Phi(y,y)}{2} - \frac{\frac{1}{2}e^{R/2}\Xi(y,y')}{\sqrt{1-\left(1-\Xi(y,y')\right)^2}}(1+\varphi_n(y,y')) \right| \\
&\qquad \le \frac{\Phi(y,y')\Xi(y,y')}{2} + \frac{\frac{1}{2}e^{R/2}\Xi(y,y')\varphi_n(y,y')}{\sqrt{1-\left(1-\Xi(y,y')\right)^2}}
\end{aligned}$$

$$\le e^{-\delta(R-y)}\Phi(y,y') + \frac{e^{-\delta(R-y)}}{1-e^{-\delta(R-y)}} \frac{e^{R/2}e^{-\delta(R-y)}}{\sqrt{2e^{-\delta(R-y)} - e^{-2\delta(R-y)}}}$$
$$\le e^{-\delta(R-y)}\Phi(y,y') + \frac{e^{\frac{R}{2}}e^{-\frac{3}{2}\delta(R-y)}}{\left(1-e^{-\delta(R-y)}\right)^{3/2}}.$$

Plugging this into (3.11) yields

$$\left|\int_0^{(1-\delta)(R-y)} \frac{\Phi(y,y')}{2} e^{-\alpha y'}\,\mathrm{d}y' - I_2(y)\right|$$
$$\le \int_0^{(1-\delta)(R-y)} \left(e^{-\delta(R-y)}\Phi(y,y') + \frac{e^{\frac{R}{2}}e^{-\frac{3}{2}\delta(R-y)}}{\left(1-e^{-\delta(R-y)}\right)^{3/2}}\right) e^{-\alpha y'}\,\mathrm{d}y'$$
$$\le e^{-\delta(R-y)}\mu\left(\mathcal{B}_\infty(y)\right) + e^{\frac{y}{2}} \frac{e^{-(\alpha-\frac{1}{2}-(\alpha-\frac{3}{2})\delta)(R-y)}}{\alpha\left(1-e^{-\delta(R-y)}\right)^{3/2}}.$$

To finish the argument we note that $R-y>0$ for all $0<y\le(1-\varepsilon)R$ and observe that $0<\delta<1$ implies that $\alpha-\frac{1}{2}-(\alpha-\frac{3}{2})\delta = \alpha-\frac{1}{2}-(\alpha-\frac{1}{2})\delta+\delta>0$. Since $\mu\left(\mathcal{B}_\infty(y)\right) = \Theta\left(e^{\frac{y}{2}}\right)$ it then follows that

$$\lim_{n\to\infty} \sup_{0<y\le(1-\varepsilon)R} \frac{2\alpha\nu}{\pi\mu\left(\mathcal{B}_\infty(y)\right)} \left|\int_0^{(1-\delta)(R-y)} \frac{\Phi(y,y')}{2} e^{-\alpha y'}\,\mathrm{d}y' - I_2(y)\right| = 0,$$

which completes the proof. □

Proof of Lemma 3.1.1, second part: This is now only a change of notation from Lemma 3.3.5: $\mu(y) = \mu\left(\mathcal{B}_\infty(y)\right)$, $\mu_{Po,n}(y) = \mu\left(\mathcal{B}(y)\right)$ and $\mu'_{Po,n}(y) = \frac{\partial}{\partial y}\mu\left(\mathcal{B}(y)\right)$. ■

3.3.3 First moment for small degrees

Proof of Lemma 3.1.2: We will show, with some asymptotic estimates and integration by substitution, that

$$\int_0^{(1-\varepsilon)R} \mathbb{P}\left(\mathrm{Po}(\mu_{Po,n}(y)) = k_n\right)\alpha e^{-\alpha y}\,\mathrm{d}y = (1+o(1))\int_0^\infty \mathbb{P}(Po(\xi e^{\frac{z}{2}}) = k_n)\alpha e^{-\alpha z}\,\mathrm{d}z.$$

This implies the result because the last integral equals $(1+o(1))2\alpha\xi^{2\alpha}k_n^{-(2\alpha+1)} = (1+o(1))p_{k_n}$.

Define the function $z(y) = 2\ln\frac{\mu_{Po,n}(y)}{\xi}$ (note that $z(y)$ is well-defined as $\mu_{Po,n}(y)\ge 0$ and that $z(y)$ is bijective because $\mu_{Po,n}(y)$ is strictly monotone increasing and continuous, see Lemma 3.3 in [29]). By rearranging, we have that

$$\mu_{Po,n}(y) = \xi e^{\frac{z(y)}{2}}.$$

From Lemma 3.1.1, it follows that uniformly for all $0 \leq y \leq (1-\epsilon)R$, $\xi e^{\frac{y}{2}} = (1+o(1))\mu_{Po,n}(y) = (1+o(1))\xi e^{\frac{z(y)}{2}}$, and hence that

$$e^{-\alpha y} = (1+o(1))e^{-\alpha z(y)}.$$

Again by Lemma 3.1.1, $\mu'_{Po,n}(y) = (1+o(1))\frac{1}{2}\mu(y) = (1+o(1))\frac{1}{2}\mu_{Po,n}(y)$ uniformly for $0 \leq y \leq (1-\epsilon)R$, and thus,

$$z'(y) = \frac{2\mu'_{Po,n}(y)}{\mu_{Po,n}(y)} = 1 + o(1).$$

From these observations, we obtain that

$$\begin{aligned}
&\int_0^{(1-\epsilon)R} \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y} dy \\
&= (1+o(1)) \int_0^{(1-\epsilon)R} \mathbb{P}(Po(\xi e^{\frac{z(y)}{2}}) = k_n)\alpha e^{-\alpha z(y)} z'(y) dy.
\end{aligned}$$

To the integral, we can apply integration by substitution, i.e. use the new variable $z = z(y)$,

$$= \int_{z(0)}^{z((1-\epsilon)R)} \mathbb{P}(Po(\xi e^{\frac{z}{2}}) = k_n)\alpha e^{-\alpha z} dz.$$

Now, it remains to show that

$$\int_{z(0)}^{z((1-\epsilon)R)} \mathbb{P}(Po(\mu(z)) = k_n)\alpha e^{-\alpha z} dz = (1+o(1)) \int_0^{\infty} \mathbb{P}(Po(\mu(z)) = k_n)\alpha e^{-\alpha z} dz. \tag{3.12}$$

We will show that $z(0) \leq y^-_{k_n,c}$ and $y^+_{k_n,c} \leq z((1-\epsilon)R)$ (where $y^-_{k_n,c}$ and $y^+_{k_n,c}$ are defined at the beginning of Section 3.1.2) and then apply the Chernoff bound to show that the difference of the integrals in (3.12) is of smaller order than the integral on the right-hand side of (3.12).

Let $C > 2\alpha + 1$. Then, there is a $c > 0$ such that the claim of Lemma 3.3.2 holds.

Now, observe that by Lemma 3.1.1, $\mu_{Po,n}(0) \to \xi$ and $\mu_{Po,n}((1-\epsilon)R) = (1+o(1))\xi e^{\frac{(1-\epsilon)R}{2}} = \Theta(n^{1-\epsilon})$. As $k_n \to \infty$, this implies that

$$\mu_{Po,n}(0) \leq k_n - c\sqrt{k_n \ln k_n}$$

for n large enough. As $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right) = o\left(n^{1-\epsilon}\right)$ (for $\epsilon < 1 + \frac{1}{2\alpha+1}$), we have that

$$k_n + c\sqrt{k_n \ln k_n} \leq \mu_{Po,n}((1-\epsilon)R)$$

for n large enough. As the function $y \mapsto 2\ln\frac{y}{2}$ is monotone increasing, it follows that $z(0) \leq y^-_{k_n,c}$ and $y^+_{k_n,c} \leq z((1-\epsilon)R)$ for n large enough. In other words, we have that

$$[y^-_{k_n,c}, y^+_{k_n,c}] \subset [z(0), z((1-\epsilon)R)].$$

From the implications of the Chernoff bound in Lemma 3.3.2, it follows that for $z \notin [z(0), z((1-\epsilon)R)]$, the Poisson probability $\mathbb{P}(Po(\mu(z)) = k_n) = O\left(k_n^{-C}\right)$. As $\int_a^\infty \alpha e^{-\alpha z}dz < \infty$ is finite for all $a \in \mathbb{R}$, it follows that the difference of the integrals in (3.12) is

$$\int_{[z(0),z((1-\epsilon)R)]\Delta[0,\infty)} \mathbb{P}(Po(\mu(z)) = k_n)\alpha e^{-\alpha z}dz = O\left(k_n^{-C}\right).$$

On the other hand $\int_0^\infty \mathbb{P}(Po(\mu(z)) = k_n)\alpha e^{-\alpha z}dz = \Theta\left(k_n^{-(2\alpha+1)}\right)$. As $C > 2\alpha+1$, this completes the proof. ■

Proof of Lemma 3.1.3: Write $N_{Po}(k_n) = \sum_{v\in V(G_{\mathrm{Po}})} \mathbb{1}_{\{D_{G_{\mathrm{Po}}}(v)=k_n\}}$. By the Campbell-Mecke formula, the expectation of $N_{Po}(k_n)$ can be expressed as an integral over $\mathbb{P}(D_{G_{\mathrm{Po}}}(v) = k_n)$ and due to the coupling Lemma 1.6.1, we can use the intensity measure μ_n with intensity function f_n:

$$\mathbb{E}\left[N_{Po}(k_n)\right] = \int_{\mathcal{R}} \mathbb{P}(D_{G_{\mathrm{Po}}}(v) = k_n)\mu_n(dv). \tag{3.13}$$

In the next step, we use that $\mathbb{P}(D_{G_{\mathrm{Po}}}(v) = k_n)$ is invariant under translations in the first coordinate,

$$\begin{aligned} &= \int_0^R \mathbb{P}(D_{G_{\mathrm{Po}}}(0,y) = k_n)\frac{\pi n}{\nu}\frac{\alpha\nu}{\pi}e^{-\alpha y}dy \\ &= n\int_0^R \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y}dy \\ &= n\left(\int_0^{(1-\epsilon)R} \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y}dy\right. \\ &\qquad \left.+\int_{(1-\epsilon)R}^R \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y}dy\right). \end{aligned} \tag{3.14}$$

Note that

$$\begin{aligned} \int_{(1-\epsilon)R}^R \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y}dy &\leq \int_{(1-\epsilon)R}^R \alpha e^{-\alpha y}dy \\ &= \Theta(e^{-\alpha(1-\epsilon)R}) = \Theta(n^{-2\alpha(1-\epsilon)}). \end{aligned}$$

Now first consider the case where $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$. Then, for $\alpha > \frac{1}{2}$ and $\epsilon > 0$ small enough, we have $2\alpha(1-\epsilon) > 1$ and hence, $\frac{2\alpha(1-\epsilon)}{2\alpha+1} > \frac{1}{2\alpha+1}$ and so,

$k_n = O\left(n^{\frac{1}{2\alpha+1}}\right) = o\left(n^{\frac{2\alpha(1-\epsilon)}{2\alpha+1}}\right)$. Taking both sides to the $-(2\alpha+1)$ implies that $k_n^{-(2\alpha+1)} \gg n^{-2\alpha(1-\epsilon)}$ or in other words, $\Theta(n^{-2\alpha(1-\epsilon)}) = o\left(k_n^{-(2\alpha+1)}\right)$. The first statement of the lemma now follows by invoking Lemma 3.1.2.

When $k_n \gg n^{\frac{1}{2\alpha+1}}$, Lemma 3.1.2 implies that

$$n\int_0^{(1-\epsilon)R} \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y}dy = (1+o\,(1))np_{k_n} = O\left(nk^{-(2\alpha+1)}\right) = o\,(1)\,.$$

On the other hand,

$$n\int_{(1-\epsilon)R}^{R} \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y}dy = O\left(n^{1-2\alpha(1-\epsilon)}\right),$$

which is $o\,(1)$ since by our choice $2\alpha(1-\epsilon) > 1$. Thus the second claim of the lemma follows.

■

3.3.4 Factorial moments for small degrees

Proof of Lemma 3.1.4: First of all, as $y^+_{k_n,c} \leq (1-\epsilon)R$, we can apply the same substitution as in the proof of Lemma 3.1.3,

$$\begin{aligned}
&\int_{\mathcal{S}_{k_n,c}} \mathbb{P}(v_1 \text{ has degree } k_n \text{ in } G_{\mathrm{Po}} \cup \{v_1\})\mu_n(dv_1) \\
&= n\int_{y^-_{k_n,c}}^{y^+_{k_n,c}} \mathbb{P}(Po(\mu_{Po,n}(y)) = k_n)\alpha e^{-\alpha y}dy \\
&= (1+o(1))n\int_{z(y^-_{k_n,c})}^{z(y^+_{k_n,c})} \mathbb{P}(Po(\mu(z)) = k_n)\alpha e^{-\alpha z}dz.
\end{aligned}$$

Let $C > 2\alpha+1$. By the implications of the Chernoff bound (see Lemma 3.3.2), there is $c > 0$ such that for $z = z(y) \notin [z(y^-_{k_n,c}), z(y^+_{k_n,c})]$,

$$\mathbb{P}(Po(\mu(z)) = k_n) = O\left(k_n^{-C}\right).$$

It follows that

$$\int_{[0,\infty)\setminus[y^-_{k_n,c},y^+_{k_n,c}]} \mathbb{P}(Po(\mu(y)) = k_n) = O\left(k_n^{-C}\right).$$

Hence, we can continue with our initial equation, and conclude the claim of the lemma by computing the resulting improper integral involving the probability mass function of the Poisson distribution

$$\int_{\mathcal{S}_{k_n,c}} \mathbb{P}(v_1 \text{ has degree } k_n \text{ in } G_{\mathrm{Po}} \cup \{v_1\})\mu_n(dv_1)$$

$$= (1+o(1))n\int_0^\infty \mathbb{P}(Po(\mu(z)) = k_n)\alpha e^{-\alpha z}dz = (1+o(1))2\alpha\xi^{2\alpha}nk_n^{-(2\alpha+1)}.$$

■

Proof of Lemma 3.1.5:

Let $H = G_{\mathrm{Po}} \cup \{v_1, \ldots, v_s\}$.

For $1 \le j \le r$, let Y_j be the number of vertices of H that are adjacent to both v_j and v_{r+1}. Let X_j be the number of vertices of H that are adjacent to v_j, but not to v_{r+1}. Then, $X_j + Y_j = D_H(v_j)$ is the degree of v_j in H.

Let X_{r+1} be the number of vertices of H that are adjacent to v_{r+1}, but to none of $v_1, \ldots, v_r$. Let Y_{r+1} be the number of vertices of H that are adjacent to v_{r+1}, and at least one of $v_1, \ldots, v_r$. Then, $X_{r+1} + Y_{r+1} = D_H(v_{r+1})$ is the degree of v_{r+1} in H.

By definition, we therefore have

$$\begin{aligned}
\varphi(\{v_1, \ldots, v_{r+1}\}; v_1, \ldots, v_s) &= \mathbb{P}(X_1 + Y_1 = \cdots = X_{r+1} + Y_{r+1} = k_n),\\
\varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s) &= \mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n),\\
\varphi(\{v_{r+1}\}; v_1, \ldots, v_s) &= \mathbb{P}(X_{r+1} + Y_{r+1} = k_n),
\end{aligned}$$

and the claim is that

$$\begin{aligned}
&\mathbb{P}(X_1 + Y_1 = \cdots = X_{r+1} + Y_{r+1} = k_n)\\
&= (1+o(1))\mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n)\mathbb{P}(X_{r+1} + Y_{r+1} = k_n) + O\left(k_n^{-C}\right).
\end{aligned}$$

Let $\epsilon' = \min(\epsilon, \epsilon(2\alpha - 1)) \in (0,1)$. As for $1 \le i \le r$, it is given that $|x_{v_i} - x_{v_{r+1}}| \ge k_n^{1+\epsilon}$, it then follows that $\mathbb{E}[Y_j] = O\left(k_n^{1-\epsilon'}\right)$ (we will see this step in more detail in Lemma 4.6.7). As $Y_{r+1} \le Y_1 + \cdots + Y_r$, we hence also have $\mu := \mathbb{E}\left[Y_{r+1}\right] = O\left(k_n^{1-\epsilon'}\right)$. In particular, there is $c_0 > 0$ such that $c_0\sqrt{k_n^{1-\epsilon'} \ln k_n^{1-\epsilon'}} \ge c_1\sqrt{\mu \ln \mu}$ (where $c_1 = \sqrt{\frac{2C}{1-\epsilon'}}$, which is well-defined because $1 - \epsilon' > 0$). Now define

$$A_n = [\mu - c_0\sqrt{k_n^{1-\epsilon'} \ln k_n^{1-\epsilon'}}, \mu + c_0\sqrt{k_n^{1-\epsilon'} \ln k_n^{1-\epsilon'}}] \cap \mathbb{N}_0.$$

By (1.13), we have

$$\begin{aligned}
\mathbb{P}(Y_{r+1} \notin A_n) &= \mathbb{P}\left(|Y_{r+1} - \mu| \ge c_0\sqrt{k_n^{1-\epsilon'} \ln k_n^{1-\epsilon'}}\right) \le \mathbb{P}\left(|Y_{r+1} - \mu| \ge c_1\sqrt{\mu \ln \mu}\right)\\
&= O\left(k_n^{-\frac{(1-\epsilon')c_1^2}{2}}\right).
\end{aligned}$$

As by definition c_1 satisfies $\frac{(1-\epsilon')c_1^2}{2} = C$, this says that for the event $S_n = \{Y_{r+1} \in A_n\}$,

$$\mathbb{P}(S_n^c) = O\left(k_n^{-C}\right).$$

Beginning with the left-hand side of the claim of the lemma, the law of total probability applied to the events $\{Y_{r+1} = y_{r+1}\}$, for all $y_{r+1} \in A_n$, and S_n^c implies that

$$\begin{aligned}
&\mathbb{P}(X_1 + Y_1 = \cdots = X_{r+1} + Y_{r+1} = k_n) \\
&= \sum_{y_{r+1} \in A_n} \mathbb{P}(X_1 + Y_1 = \cdots = X_{r+1} + y_{r+1} = k_n | Y_{r+1} = y_{r+1}) \mathbb{P}(Y_{r+1} = y_{r+1}) \\
&\qquad + \mathbb{P}(X_1 + Y_1 = \cdots = X_{r+1} + Y_{r+1} = k_n | S_n^c) \mathbb{P}(S_n^c) \\
&= \sum_{y_{r+1} \in A_n} \mathbb{P}(X_1 + Y_1 = \cdots = X_{r+1} + y_{r+1} = k_n | Y_{r+1} = y_{r+1}) \mathbb{P}(Y_{r+1} = y_{r+1}) \\
&\qquad + O\left(k_n^{-C}\right).
\end{aligned}$$

As X_{r+1} is independent of $X_1, \ldots, X_r, Y_1, \ldots, Y_r$ by the properties of a Poisson process (as X_{r+1} counts the number of points in a set which is disjoint to what $X_1, \ldots, X_r, Y_1, \ldots, Y_r$ are counting), it follows that

$$\begin{aligned}
= \sum_{y_{r+1} \in A_n} &\mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n | Y_{r+1} = y_{r+1}) \\
&\times \mathbb{P}(X_{r+1} + y_{r+1} = k_n) \mathbb{P}(Y_{r+1} = y_{r+1}) + O\left(k_n^{-C}\right).
\end{aligned} \tag{3.15}$$

We will now show that uniformly for all $y_{r+1}, s \in A_n$, it holds that,

$$\mathbb{P}(X_{r+1} = k_n - y_{r+1}) = (1 + o(1)) \mathbb{P}(X_{r+1} = k_n - s). \tag{3.16}$$

To see this, observe that for all $y_{r+1}, s \in A_n$, we have that

$$|y_{r+1} - s| \leq 2c\sqrt{k_n^{1-\epsilon'} \ln k_n^{1-\epsilon'}}.$$

Denote the expectation of X_{r+1} by λ, write $\delta_n = k_n - y_{r+1} - \lambda$ and note that

$$\begin{aligned}
\frac{\mathbb{P}(X_{r+1} = k_n - y_{r+1})}{\mathbb{P}(X_{r+1} = k_n - s)} &= \frac{\mathbb{P}(X_{r+1} = k_n - y_{r+1})}{\mathbb{P}(X_{r+1} = k_n - y_{r+1} + (y_{r+1} - s))} \\
&= \frac{(k_n - y_{r+1} + (y_{r+1} - s))!}{(k_n - y_{r+1})!} \lambda^{s - y_{r+1}}.
\end{aligned}$$

We will now use that $\frac{(a+b)!}{a!} = (1 + o(1))(a + b)^b$ for $b^2 = o(a)$, applied to $a = k_n - y_{r+1}$ and $b = y_{r+1} - s$. To see this auxiliary fact, note that by Stirling's approximation to the factorial, it follows that

$$\frac{(a+b)!}{a!} = (1 + o(1)) \frac{(a+b)^{a+b+\frac{1}{2}} e^{-a-b}}{a^{a+\frac{1}{2}} e^{-a}} = \left(1 + \frac{b}{a}\right)^{a+\frac{1}{2}} (a+b)^b e^{-b}.$$

We have that $1 + \frac{b}{a} \leq e^{\frac{b}{a}}$, which implies $(1 + \frac{b}{a})^a \leq e^b$. Furthermore, as $\ln(1+x) \geq x - \frac{x^2}{2}$, we have that $(1 + \frac{b}{a})^a = e^{a \ln(1 + \frac{b}{a})} \geq e^{a(\frac{b}{a} - \frac{b^2}{2a})} = e^{b - \frac{b^2}{2a}} = (1 + o(1)) e^b$

because $b^2 = o(a)$. Finally, $b^2 = o(a)$ also implies that $(1+\frac{b}{a})^{\frac{1}{2}} = 1 + o(1)$ (here, $b = o(a)$ would have been sufficient already). This finishes the proof of the auxiliary fact and we can continue with

$$\begin{aligned}
\frac{\mathbb{P}(X_{r+1} = k_n - y_{r+1})}{\mathbb{P}(X_{r+1} = k_n - s)} &= (1+o(1))(k_n - y_{r+1} + (y_{r+1} - s))^{y_{r+1}-s}\lambda^{s-y_{r+1}} \\
&= (1+o(1))(\lambda + \delta_n + (y_{r+1} - s))^{y_{r+1}-s}\lambda^{s-y_{r+1}} \\
&= (1+o(1))\left(1 + \frac{\delta_n + (y_{r+1}-s)}{\lambda}\right)^{y_{r+1}-s} \\
&= (1+o(1))e^{\frac{\delta_n(y_{r+1}-s)}{\lambda}}e^{\frac{(y_{r+1}-s)^2}{\lambda}} = 1 + o(1),
\end{aligned}$$

where the last line follows since $\delta_n, |y_{r+1} - s| \le 2c_0\sqrt{k_n^{1-\epsilon'}\ln k_n^{1-\epsilon'}}$ and $\lambda = \Theta(k_n)$ and therefore, with convergence uniform in y_{r+1}, s,

$$\frac{\delta_n(y_{r+1}-s)}{\lambda}, \frac{(y_{r+1}-s)^2}{\lambda} \le \frac{4c_0^2 k_n^{1-\epsilon'}\ln k_n^{1-\epsilon'}}{\lambda} \to 0.$$

As $\mathbb{P}(S_n^c) = O\left(k_n^{-C}\right)$, we have

$$1 = \sum_{s\in A_n}\mathbb{P}(Y_{r+1} = s) + O\left(k_n^{-C}\right).$$

From (3.16), it follows that

$$\begin{aligned}
\mathbb{P}(X_{r+1} = k_n - y_{r+1}) &= \mathbb{P}(X_{r+1} = k_n - y_{r+1})\sum_{s\in A_n}\mathbb{P}(Y_{r+1} = s) + O\left(k_n^{-C}\right) \\
&= (1+o(1))\sum_{s\in A_n}\mathbb{P}(Y_{r+1} = s)\mathbb{P}(X_{r+1} = k_n - s) + O\left(k_n^{-C}\right) \\
&= (1+o(1))\sum_{s\in A_n}\mathbb{P}(X_{r+1} + Y_{r+1} = k_n, Y_{r+1} = s) + O\left(k_n^{-C}\right) \\
&= (1+o(1))\mathbb{P}(X_{r+1} + Y_{r+1} = k_n, Y_{r+1} \in A_n) + O\left(k_n^{-C}\right) \\
&= (1+o(1))\left(\mathbb{P}(X_{r+1} + Y_{r+1} = k_n)\right. \\
&\qquad \left.-\mathbb{P}(X_{r+1} + Y_{r+1} = k_n, Y_{r+1} \notin A_n)\right) + O\left(k_n^{-C}\right) \\
&= (1+o(1))\mathbb{P}(X_{r+1} + Y_{r+1} = k_n) + O\left(k_n^{-C}\right).
\end{aligned}$$

We insert this into (3.15). We note that the $O\left(k_n^{-C}\right)$ error term can be taken outside of the summation (which is bounded by 1) and that $\mathbb{P}(X_{r+1} + Y_{r+1} = k_n)$ does not depend on y_{r+1}, so can be put in front of the summation. Overall, this yields that

$$\sum_{y_{r+1}\in A_n}\mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n | Y_{r+1} = y_{r+1})$$

$$
\begin{aligned}
&\quad \times \mathbb{P}(X_{r+1} + y_{r+1} = k_n)\mathbb{P}(Y_{r+1} = y_{r+1}) + O\left(k_n^{-C}\right) \\
&= (1+o(1))\mathbb{P}(X_{r+1} + Y_{r+1} = k_n) \\
&\quad \times \sum_{y_{r+1}\in A_n} \mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n | Y_{r+1} = y_{r+1})\mathbb{P}(Y_{r+1} = y_{r+1}) \\
&\quad + O\left(k_n^{-C}\right).
\end{aligned}
$$

Furthermore, we have that

$$
\begin{aligned}
&\sum_{y_{r+1}\in A_n} \mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n | Y_{r+1} = y_{r+1})\mathbb{P}(Y_{r+1} = y_{r+1}) \\
&= \mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n, Y_{r+1} \in A_n) \\
&= (1+o(1))\mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n) + O\left(k_n^{-C}\right).
\end{aligned}
$$

Plugging this into the previous step finally gives the right-hand side of the claim to show the lemma,

$$
\begin{aligned}
&= (1+o(1))\mathbb{P}(X_{r+1} + Y_{r+1} = k_n) \\
&\quad \times \sum_{y_{r+1}\in A_n} \mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n | Y_{r+1} = y_{r+1})\mathbb{P}(Y_{r+1} = y_{r+1}) \\
&\quad + O\left(k_n^{-C}\right) \\
&= (1+o(1))\mathbb{P}(X_{r+1} + Y_{r+1} = k_n)\mathbb{P}(X_1 + Y_1 = \cdots = X_r + Y_r = k_n) + O\left(k_n^{-C}\right).
\end{aligned}
$$

■

Proof of Lemma 3.1.6: Let $c > 0$ be such that the claim of Lemma 3.1.4 holds.

For every fixed positive integer s, we will show by induction on r, $1 \le r \le s$, that for all $v_{r+1}, \ldots, v_s \in \mathcal{S}_{k_n,c}$, it holds that

$$
\begin{aligned}
&\int_{\mathcal{S}_{k_n,c}} \cdots \int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s)\mu_n(dv_1)\cdots\mu_n(dv_r) \\
&= (1+o(1))\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)^r.
\end{aligned}
$$

Note that $r = s$ is the claim of the lemma.

For $r = 1$, we only need to show that uniformly for $v_1 \in \mathcal{S}_{k_n,c}$,

$$
\varphi(\{v_1\}; v_1, \ldots, v_s) = (1+o(1))\varphi(\{v_1\}; v_1).
$$

To see this, note that as $v_1 \in \mathcal{S}_{k_n,c}$, the expected degree of v_1 in G_{Po} is $\Theta(k_n)$. Assume that v_1 is adjacent to $s' < s$ many vertices among $v_2, \ldots, v_s$. Then, as s' is finite and $k_n \to \infty$, we have that

$$
\mathbb{P}(D_{G_{\mathrm{Po}}}(v_1) = k_n - s') = (1+o(1))\mathbb{P}(D_{G_{\mathrm{Po}}}(v_1) = k_n).
$$

So, we have the base case of the induction,

$$\begin{aligned}\varphi(\{v_1\}; v_1, \ldots, v_s) &= \mathbb{P}(D_H(v_1) = k_n) = \mathbb{P}(D_{G_{\mathrm{Po}}}(v_1) = k_n - s') \\ &= (1 + o(1))\mathbb{P}(D_{G_{\mathrm{Po}}}(v_1) = k_n) = (1 + o(1))\varphi(\{v_1\}; v_1).\end{aligned}$$

Assuming the claim holds for integer $r < s$, we will show that it holds for $r + 1$.

Let $v_{r+2}, \ldots, v_s \in \mathcal{S}_{k_n,c}$ (if $r + 2 > s$, this definition is void and the corresponding points will never be used in the proof). Fix $0 < \epsilon < 1$ small enough s.t. $\frac{1}{2} + \epsilon < \alpha$. Define the region that the $(r + 1)$-th vertex v_{r+1} is far apart from all other vertices horizontally,

$$\mathcal{F}_\epsilon(k_n) = \{(v_1, \ldots, v_{r+1}) \in (\mathcal{S}_{k_n,c})^{r+1} : \forall 1 \le i \le r : |x_{v_i} - x_{v_{r+1}}| \ge k_n^{1+\epsilon}\}.$$

We will split the integration into this region and its complement $\mathcal{F}_\epsilon(k_n)^c = (\mathcal{S}_{k_n,c})^{r+1} \backslash \mathcal{F}_\epsilon(k_n)$.

First of all, we derive an upper bound for the complement $\mathcal{F}_\epsilon(k_n)^c$. Note that

$$\varphi(\{v_1, \ldots, v_{r+1}\}; v_1, \ldots, v_s) \le \varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s)$$

and so,

$$\begin{aligned}&\int\int_{\mathcal{F}_\epsilon(k_n)^c} \varphi(\{v_1, \ldots, v_{r+1}\}; v_1, \ldots, v_s)\mu_n(dv_1)\ldots\mu_n(dv_{r+1}) \\ &\le \int\int_{\mathcal{F}_\epsilon(k_n)^c} \varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s)\mu_n(dv_1)\ldots\mu_n(dv_{r+1}).\end{aligned}$$

For $(v_1, \ldots, v_{r+1}) \in \mathcal{F}_\epsilon(k_n)^c$, we have that $(v_1, \ldots, v_r) \in (\mathcal{S}_{k_n,c})^r$ and $v_{r+1} = (x_{r+1}, y_{r+1})$ satisfies $y^-_{k_n,c} \le y_{r+1} \le y^+_{k_n,c}$ and x_{r+1} falls into an interval I_n of length $2k_n^{1+\epsilon}$. As the integrand $\varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s)$ is constant in x_{r+1}, we can upper bound the corresponding integration w.r.t. $\mu_n(dv_{r+1})$ resp. dy_{r+1} as follows,

$$\begin{aligned}&\int_{\{(x_{r+1},y_{r+1})\in\mathcal{S}_{k_n,c}:x_{r+1}\in I_n\}} \varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s)\mu_n(dv_{r+1}) \\ &= \int_{y^-_{k_n,c}}^{y^+_{k_n,c}} \int_{I_n} \varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s)\frac{\alpha\nu}{\pi}e^{-\alpha y_{r+1}}dx_{r+1}dy_{r+1} \\ &\le 2k_n^{1+\epsilon}\varphi(\{v_1, \ldots, v_r\}; v_1, \ldots, v_s)\int_{y^-_{k_n,c}}^{y^+_{k_n,c}} \frac{\alpha\nu}{\pi}e^{-\alpha y_{r+1}}dy_{r+1}.\end{aligned}$$

Furthermore, we have that

$$\int_{y^-_{k_n,c}}^{y^+_{k_n,c}} e^{-\alpha y_1}dy_1 \le \int_{y^-_{k_n,c}}^{\infty} e^{-\alpha y_1}dy_1 = O\left(e^{-\alpha y^-_{k_n,c}}\right)$$

$$= O\left(\left(\frac{k_n - c\sqrt{k_n \ln k_n}}{\xi}\right)^{-2\alpha}\right) = O\left(k_n^{-2\alpha}\right).$$

We have thus established that

$$\int\int_{\mathcal{F}_\epsilon(k_n)^c} \varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_s)\mu_n(dv_1)\dots\mu_n(dv_{r+1})$$
$$\leq O\left(k_n^{1+\epsilon}k_n^{-2\alpha}\right)\int_{\mathcal{S}_{k_n,c}}\cdots\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_s)\mu_n(dv_1)\dots\mu_n(dv_r).$$

By applying the induction hypothesis to the r-fold integral, we obtain the upper bound

$$= O\left(k_n^{1+\epsilon-2\alpha}\right)\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)^r. \tag{3.17}$$

Finally, we use that $k_n^{1+\epsilon-2\alpha} = o\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)$. To see this, we will firstly show that $k_n^{1+\epsilon-2\alpha} = o\left(nk_n^{-(2\alpha+1)}\right)$ and then apply Lemma 3.1.4, which says that

$$\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1) = \Theta\left(nk_n^{-(2\alpha+1)}\right).$$

By our choice of ϵ, we have that $\frac{1}{2} + \epsilon < \alpha$, which implies that $\frac{2+\epsilon}{1+2\alpha} < 1$, and hence, using $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$,

$$\frac{k_n^{1+\epsilon-2\alpha}}{nk_n^{-(2\alpha+1)}} = n^{-1}k_n^{2+\epsilon} = O\left(n^{-1}n^{\frac{2+\epsilon}{1+2\alpha}}\right) = o\left(n^{-1}n\right) = o\left(1\right).$$

Having shown that $k_n^{1+\epsilon-2\alpha} = o\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)$, we therefore conclude from (3.17) with the upper bound

$$= o\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)^r$$
$$= o\left(\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\}; v_1)\mu_n(dv_1)\right)^{r+1}\right).$$

For the integration over $\mathcal{F}_\epsilon(k_n)$, recall that by Lemma 3.1.5,

$$\varphi(\{v_1, \dots, v_{r+1}\}; v_1, \dots, v_s)$$
$$= (1 + o(1))\varphi(\{v_1, \dots, v_r\}; v_1, \dots, v_s)\varphi(\{v_{r+1}\}; v_1, \dots, v_s) + O\left(k_n^{-C}\right).$$

This implies that

$$\begin{aligned}
&\int\int_{\mathcal{F}_\epsilon(k_n)} \varphi(\{v_1,\ldots,v_{r+1}\};v_1,\ldots,v_s)\mu_n(dv_1)\ldots\mu_n(dv_{r+1})\\
&= (1+o(1))\int\int_{\mathcal{F}_\epsilon(k_n)} \varphi(\{v_1,\ldots,v_r\};v_1,\ldots,v_s)\\
&\qquad\qquad \times\varphi(\{v_{r+1}\};v_1,\ldots,v_s)\mu_n(dv_1)\ldots\mu_n(dv_{r+1})\\
&\qquad + O\left(k_n^{-C}\right)\int\int_{\mathcal{F}_\epsilon(k_n)} \mu_n(dv_1)\ldots\mu_n(dv_{r+1}) =: M+E,
\end{aligned}$$

where

$$\begin{aligned}
M &= (1+o(1))\int\int_{\mathcal{F}_\epsilon(k_n)} \varphi(\{v_1,\ldots,v_r\};v_1,\ldots,v_s)\\
&\qquad\qquad \times\varphi(\{v_{r+1}\};v_1,\ldots,v_s)\mu_n(dv_1)\ldots\mu_n(dv_{r+1}),\\
E &= O\left(k_n^{-C}\right)\int\int_{\mathcal{F}_\epsilon(k_n)} \mu_n(dv_1)\ldots\mu_n(dv_{r+1}).
\end{aligned}$$

We will show that

$$M = (1+o(1))\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\};v_1)\mu_n(dv_1)\right)^{r+1}$$

and

$$E = o\left(\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\};v_1)\mu_n(dv_1)\right)^{r+1}\right),$$

after which the proof will be finished.

Regarding M, note that we can factorize into an integration over $v_1,\ldots,v_r$ and one over v_{r+1}.

Furthermore, we note that the condition $(v_1,\ldots,v_{r+1})\in\mathcal{F}_\epsilon(k_n)$ implies that (writing $v_{r+1}=(x_{r+1},y_{r+1})$) the horizontal coordinate x_{r+1} falls into an interval I_n of length L_n, satisfying

$$\frac{\pi n}{\nu} - 2rk_n^{1+\epsilon} \le L_n \le \frac{\pi n}{\nu} - 2k_n^{1+\epsilon}.$$

As $k_n^{1+\epsilon} = O\left(n^{\frac{1+\epsilon}{2\alpha+1}}\right) = o(n)$ for $\epsilon<1$ and $\alpha>\frac{1}{2}$, this shows that the length of the integration range in x_{r+1} satisfies $L_n=(1+o(1))\frac{\pi n}{\nu}$. Thus, we have that

$$M = (1+o(1))\int_{I_n}\int_{y^-_{k_n,c}}^{y^+_{k_n,c}} \varphi(\{v_{r+1}\};v_1,\ldots,v_s)\frac{\alpha\nu}{\pi}e^{-\alpha y_{r+1}}dy_{r+1}dx_{r+1}$$

$$
\begin{aligned}
&\times \int_{\mathcal{S}_{k_n,c}} \cdots \int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1,\ldots,v_r\};v_1,\ldots,v_s)\mu_n(dv_1)\ldots\mu_n(dv_r) \\
&= (1+o(1))n\int_{y^-_{k_n,c}}^{y^+_{k_n,c}} \varphi(\{v_{r+1}\};v_1,\ldots,v_s)\alpha e^{-\alpha y_{r+1}}dy_{r+1} \\
&\qquad \times \int_{\mathcal{S}_{k_n,c}} \cdots \int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1,\ldots,v_r\};v_1,\ldots,v_s)\mu_n(dv_1)\ldots\mu_n(dv_r) \\
&= (1+o(1))\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_{r+1}\};v_1,\ldots,v_s)\mu_n(dv_1) \\
&\qquad \times \int_{\mathcal{S}_{k_n,c}} \cdots \int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1,\ldots,v_r\};v_1,\ldots,v_s)\mu_n(dv_1)\ldots\mu_n(dv_r).
\end{aligned}
$$

By applying the base case of the induction to the first factor and the induction hypothesis to the second one, we have derived that

$$
M = (1+o(1))\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\};v_1)\mu_n(dv_1)\right)^{r+1}.
$$

Regarding E, we observe that

$$
E = O\left(k_n^{-C}\right)\int\int_{\mathcal{F}_\epsilon(k_n)} \mu_n(dv_1)\ldots\mu_n(dv_{r+1}) = O\left(k_n^{-C}\right)(1+o(1))n^{r+1}.
$$

Recall that again by Lemma 3.1.4,

$$
\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\};v_1)\mu_n(dv_1) = \Theta\left(nk_n^{-(2\alpha+1)}\right),
$$

which implies that

$$
\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\};v_1)\mu_n(dv_1)\right)^{r+1} = \Theta\left(n^{r+1}k_n^{-r(2\alpha+1)}\right).
$$

For $C > r(2\alpha+1)$, we can conclude that

$$
E = O\left(n^{r+1}k_n^{-C}\right) = o\left(n^{r+1}k_n^{-r(2\alpha+1)}\right) = o\left(\left(\int_{\mathcal{S}_{k_n,c}} \varphi(\{v_1\};v_1)\mu_n(dv_1)\right)^{r+1}\right).
$$

■

Proof of Lemma 3.1.7:

Let $C > r(2\alpha+1)$ and take $c > 0$ from Lemma 3.3.2. For $(v_1,\ldots,v_r) \in (\mathcal{R}\times\cdots\times\mathcal{R})\backslash(\mathcal{S}_{k_n,c}\times\cdots\times\mathcal{S}_{k_n,c})$, there is a j, $1 \le j \le r$, such that $y_j \notin$

$[y^-_{k_n,c}, y^+_{k_n,c}]$, so the implications of the Chernoff bound (as in Lemma 3.3.2) yield that $\mathbb{P}(D_{G_{\mathrm{Po}}}(v_j) = k_n) = O\left(k_n^{-C}\right)$. As, for $1 \leq j \leq r$, the event $\{D_{G_{\mathrm{Po}}}(v_1) = \cdots = D_{G_{\mathrm{Po}}}(v_r) = k_n\}$ implies the event $\{D_{G_{\mathrm{Po}}}(v_j) = k_n\}$, it follows that

$$\mathbb{P}(D_{G_{\mathrm{Po}}}(v_1) = \cdots = D_{G_{\mathrm{Po}}}(v_r) = k_n) = O\left(k_n^{-C}\right)$$

and hence,

$$\begin{aligned}\int\cdots\int_{(\mathcal{R}\times\cdots\times\mathcal{R})\setminus(\mathcal{S}_{k_n,c}\times\cdots\times\mathcal{S}_{k_n,c})} &\mathbb{P}(D_{G_{\mathrm{Po}}}(v_1) = \cdots = D_{G_{\mathrm{Po}}}(v_r) = k_n)\mu_n(dv_1)\cdots\mu_n(dv_r)\\ &= O\left(n^r k_n^{-C}\right).\end{aligned}$$

By Lemma 3.1.6 and Lemma 3.1.4,

$$\begin{aligned}&\int_{\mathcal{S}_{k_n,c}}\cdots\int_{\mathcal{S}_{k_n,c}} \mathbb{P}(v_1,\ldots,v_r \text{ have degree } k_n \text{ in } G_{\mathrm{Po}}\cup\{v_1,\ldots,v_r\})\mu_n(dv_1)\cdots\mu_n(dv_r)\\ &= (1+o(1))\left(\int_{\mathcal{S}_{k_n,c}} \mathbb{P}(v_1 \text{ has degree } k_n \text{ in } G_{\mathrm{Po}}\cup\{v_1\})\mu_n(dv_1)\right)^r\\ &= (1+o(1))\left(2\alpha\xi^{2\alpha} n k_n^{-(2\alpha+1)}\right)^r.\end{aligned}$$

Hence, the claim follows as $C > r(2\alpha+1)$. ■

Proof of Lemma 3.1.8: First apply Lemma 3.1.7 to the left-hand side of the claim, then apply Lemma 3.1.6 and finally apply Lemma 3.1.7 again, in the reverse direction for $r = 1$. ■

Proof of Lemma 3.1.9: First of all, we observe that

$$\binom{N_{Po}(k_n)}{r} = \frac{1}{r!}\sum_{v_1,\ldots,v_r\in V(G_{\mathrm{Po}})}^{\neq} \mathbb{1}_{\{D_{G_{\mathrm{Po}}}(v_1)=\cdots=D_{G_{\mathrm{Po}}}(v_r)=k_n\}}.$$

This can be seen by induction on r. For $r = 1$, the claim is clear. Assuming it holds for $r \geq 1$, by the induction hypothesis,

$$\begin{aligned}\binom{N_{Po}(k_n)}{r+1} &= \binom{N_{Po}(k_n)}{r}\frac{N_{Po}(k_n)-r}{r+1}\\ &= \frac{1}{(r+1)!}\sum_{v_1,\ldots,v_r\in V(G_{\mathrm{Po}})}^{\neq} \mathbb{1}_{\{D_{G_{\mathrm{Po}}}(v_1)=\cdots=D_{G_{\mathrm{Po}}}(v_r)=k_n\}}(N_{Po}(k_n)-r).\end{aligned}$$

Now, we can write

$$N_{Po}(k_n) = \sum_{\substack{v_{r+1}\in V(G_{\mathrm{Po}}),\\ v_{r+1}\notin\{v_1,\ldots,v_r\}}} \mathbb{1}_{\{D_{G_{\mathrm{Po}}}(v_{r+1})=k_n\}} + \sum_{v_{r+1}\in\{v_1,\ldots,v_r\}} \mathbb{1}_{\{D_{G_{\mathrm{Po}}}(v_{r+1})=k_n\}}.$$

The first sum leads to the right-hand side of the claim for $r+1$, whereas the second sum will cancel with the $-r$.

Recall that due to Lemma 1.6.1, we can use the intensity measure μ_n with intensity function $f_n = f_n(x,y) = \frac{\alpha\nu}{\pi}e^{-\alpha y}\mathbb{1}_{\{-\frac{\pi n}{2\nu}<x\leq\frac{\pi n}{2\nu},0<y<2\ln\frac{n}{\nu}\}}$ for the Poisson point process representing the vertex set of G_{Po}. By the Campbell-Mecke formula (1.11),

$$\mathbb{E}\left[\binom{N_{Po}(k_n)}{r}\right] = \frac{1}{r!}\mathbb{E}\left[\sum_{v_1,\ldots,v_r\in V(G_{\mathrm{Po}})}^{\neq}\mathbb{1}_{\{D_{G_{\mathrm{Po}}}(v_1)=\cdots=D_{G_{\mathrm{Po}}}(v_r)=k_n\}}\right]$$
$$= \frac{1}{r!}\int_{\mathcal{R}}\cdots\int_{\mathcal{R}}\mathbb{P}(v_1,\ldots,v_r \text{ have degree } k_n \text{ in } G_{\mathrm{Po}}\cup\{v_1,\ldots,v_r\})\mu_n(dv_1)\cdots\mu_n(dv_r),$$

where we integrate over r additional points which we can think of as being added independently and with the same distribution as the vertices of the poissonized KPKVB model G_{Po} in the upper half-plane coordinates.

With $r=1$, it follows that

$$\mathbb{E}\left[N_{Po}(k_n)\right] = \int_{\mathcal{R}}\mathbb{P}(v_1 \text{ has degree } k_n \text{ in } G_{\mathrm{Po}}\cup\{v_1\})\mu_n(dv_1),$$

which yields that the right-hand side of the claim of the lemma can be rewritten as

$$\frac{1}{r!}(\mathbb{E}\left[N_{Po}(k_n)\right])^r = \frac{1}{r!}\left(\int_{\mathcal{R}}\mathbb{P}(v_1 \text{ has degree } k_n \text{ in } G_{\mathrm{Po}}\cup\{v_1\})\mu_n(dv_1)\right)^r.$$

Using Lemma 3.1.8, we conclude that

$$\mathbb{E}\left[\binom{N_{Po}(k_n)}{r}\right]$$
$$= \frac{1}{r!}\int\limits_{\mathcal{R}}\cdots\int\limits_{\mathcal{R}}\mathbb{P}(v_1,\ldots,v_r \text{ have degree } k_n \text{ in } G_{\mathrm{Po}}\cup\{v_1,\ldots,v_r\})\mu_n(dv_1)\cdots\mu_n(dv_r)$$
$$= (1+o(1))\frac{1}{r!}\left(\int_{\mathcal{R}}\mathbb{P}(v_1 \text{ has degree } k_n \text{ in } G_{\mathrm{Po}}\cup\{v_1\})\mu_n(dv_1)\right)^r$$
$$= (1+o(1))\frac{1}{r!}(\mathbb{E}\left[N_{Po}(k_n)\right])^r.$$

■

3.3.5 Poissonization for small degrees

Proof of Lemma 3.1.10: We consider the coupling where we have an infinite supply of i.i.d. points $u_1, u_2, \ldots$ chosen according to the (α, R)-quasi uniform distribution, the vertices of $G(n;\alpha,\nu)$ are $u_1,\ldots,u_n$ and the vertices of $G_{\mathrm{Po}}(n;\alpha,\nu)$

are $u_1, \dots, u_N$ with $N \stackrel{\mathrm{d}}{=} \mathrm{Po}(n)$ independently of $u_1, u_2, \dots$. Thus, under this coupling, the only difference between $G_n = G(n; \alpha, \nu)$ and $G_{\mathrm{Po}} = G_{\mathrm{Po}}(n; \alpha, \nu)$ is the number of points. Note that since N is Poisson with expectation n, it follows from the Chernoff bound (see also Equation (1.13)) that we may assume that $n - C\sqrt{n \log n} \le N \le n + C\sqrt{n \log n}$. To keep notation simple we will suppress this conditioning in the derivations.

Clearly, if $N = n$ the graphs are the same. So we will consider the two cases $n - C\sqrt{n \log n} \le N < n$ and $n < N \le n + C\sqrt{n \log n}$. We will prove the latter case. The other case uses similar arguments and hence we omit the details here.

If $n < N \le n + C\sqrt{n \log n}$, then G_n has less vertices than G_{Po}. Write $V_n(k_n)$ and $V_{\mathrm{Po}}(k_n)$ to denote the set of nodes that have degree k_n in G_n and G_{Po}, respectively. Then since vertices $u_{n+1}, \dots, u_N$ are not present in G_n,

$$
\begin{aligned}
|N_n(k_n) - N_{\mathrm{Po}}(k_n)| &= \sum_{i=1}^{N} \mathbb{1}_{\{u_i \in V_n(k_n) \Delta V_{\mathrm{Po}}(k_n)\}} \\
&= \sum_{i=1}^{n} \mathbb{1}_{\{u_i \in V_n(k_n) \Delta V_{\mathrm{Po}}(k_n)\}} + \sum_{i=n+1}^{N} \mathbb{1}_{\{u_i \in V_{\mathrm{Po}}(k_n)\}}.
\end{aligned}
$$

Let D_{Po} denote the degree in the Poissonized KPKVB model of a point u placed according to the (α, R)-quasi uniform distribution. Then, the expectation of the second summation equals

$$
\begin{aligned}
(N-n)\mathbb{P}\left(D_{\mathrm{Po}} = k_n\right) &= (N-n) \int_0^R \mathbb{P}\left(\mathrm{Po}(\mu_{Po,n}(y)) = k_n\right) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&\le (1 + o(1)) C \sqrt{n \log n} p_{k_n} = o\left(\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]\right).
\end{aligned}
$$

Therefore, it remains to consider the first summation.

Let $D_n(u)$ and $D_{\mathrm{Po}}(u)$ denote the degree of a point u in G_n and G_{Po}, respectively. Then, there are two scenarios to consider: 1) either $D_n(u_i) = k_n$ and $D_{\mathrm{Po}}(u_i) \neq k_n$ or 2) $D_n(u_i) \neq k_n$ and $D_{\mathrm{Po}}(u_i) = k_n$. In the first case, since u_i is present in both graphs it follows that $D_{\mathrm{Po}}(u_i) > k_n$. Similarly, for the second case it must hold that $D_n(u_i) < k_n$. Hence we have

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathbb{1}_{\{u_i \in V_n(k_n) \Delta V_{\mathrm{Po}}(k_n)\}} \\
&= \sum_{i=1}^{n} \mathbb{1}_{\{d_n(u_i) = k_n, D_{\mathrm{Po}}(u_i) > k_n\}} + \sum_{i=1}^{n} \mathbb{1}_{\{d_n(u_i) < k_n, D_{\mathrm{Po}}(u_i) = k_n\}}.
\end{aligned}
$$

Let us first consider the second summation, i.e. the case where the vertex has degree smaller than k_n in G_n. Taking the expectation gives $n\mathbb{P}\left(D_n < k_n, D_{\mathrm{Po}} = k_n\right)$, where D_n denotes the degree in the KPKVB model of a point u placed according to the (α, R)-quasi uniform distribution. We now observe that because the points $u_1, \dots, u_N$ used to couple the graphs are independent, we can view the graph G_n as being obtained from G_{Po} by removing $N - n$ points, uniformly at random. Therefore, if a vertex has degree k_n in G_{Po} but smaller degree in G_n, this means that at least one of its neighbours was removed. Denote by $Z(n)$ a random variable with a Hypergeometric distribution, for taking $N-n$ draws from a population of size N, where there are k_n good objects. That is, $Z(n)$ denotes the number of removed neighbours of a vertex u with degree k_n in G_{Po}. We then have

$$\begin{aligned}\mathbb{P}\left(D_n < k_n, D_{\mathrm{Po}} = k_n\right) &= \mathbb{P}\left(Z(n) > 1\right)\mathbb{P}\left(D_{\mathrm{Po}} = k_n\right) \\ &\leq \mathbb{E}\left[Z(n)\right]\mathbb{P}\left(D_{\mathrm{Po}} = k_n\right) = \frac{(N-n)k_n}{N}\mathbb{P}\left(D_{\mathrm{Po}} = k_n\right).\end{aligned}$$

Because $\alpha > 1/2$ and $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$, it holds that $k_n = o\left(\sqrt{n}\right)$. Since $N = \Theta\left(n\right)$ and $N - n \leq O\left(\sqrt{n \log n}\right)$, it then follows that $\frac{(N-n)k_n}{N} = o\left(1\right)$, from which we conclude that

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\{d_n(u_i) < k_n, D_{\mathrm{Po}}(u_i) = k_n\}}\right] &= n\mathbb{P}\left(D_n < k_n, D_{\mathrm{Po}} = k_n\right) \\ &\leq o\left(1\right) n\mathbb{P}\left(D_{\mathrm{Po}} = k_n\right) = o\left(\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]\right).\end{aligned}$$

We now proceed with the other summation, for the case where a vertex has degree k_n in G_n but larger degree in G_{Po}. Since the degree of u in G_{Po} can be at most $N - n$ larger we have

$$\mathbb{P}\left(D_n(u) = k_n, D_{\mathrm{Po}}(u) > k_n\right) = \sum_{t=1}^{N-n} \mathbb{P}\left(D_n(u) = k_n, D_{\mathrm{Po}}(u) = k_n + t\right).$$

Using that the graph G_n can be seen as being obtained from G_{Po} by removing $N - n$ points uniformly at random, a point with degree $k_n + t$ in G_{Po} can only have degree k_n in G_n if exactly t of its neighbours were removed. Let us therefore denote by $Z(n,t)$ a random variable with a Hypergeometric distribution, for taking $N - n$ draws from a population of size N, where there are $k_n + t$ good objects. Then,

$$\mathbb{P}\left(D_n(u) = k_n, D_{\mathrm{Po}}(u) = k_n + t\right) = \mathbb{P}\left(Z(n,t) = t\right)\mathbb{P}\left(D_{\mathrm{Po}} = k_n + t\right).$$

Recall that, for any $0 < \varepsilon < 1$,

$$\begin{aligned}\mathbb{P}\left(D_{\mathrm{Po}} = k_n + t\right) &= \int_0^R \mathbb{P}\left(\mathrm{Po}(\mu_{Po,n}(y)) = k_n + t\right)\alpha e^{-\alpha y}\,\mathrm{d}y \\ &= \int_0^{(1-\varepsilon)R} \mathbb{P}\left(\mathrm{Po}(\mu_{Po,n}(y)) = k_n + t\right)\alpha e^{-\alpha y}\,\mathrm{d}y\end{aligned}$$

$$+\int_{(1-\varepsilon)R}^{R} \mathbb{P}\left(\mathrm{Po}(\mu_{Po,n}(y)) = k_n + t\right) \alpha e^{-\alpha y}\,\mathrm{d}y$$

By Lemma 3.1.2 the first part is $(1+o(1))p_{k_n+t}$ while the second part is $O(n^{-2\alpha(1-\varepsilon)})$ and hence,

$$\mathbb{P}\left(D_{\mathrm{Po}} = k_n + t\right) \leq O\left(1\right)\left(p_{k_n+t} + n^{-2\alpha(1-\varepsilon)}\right).$$

In addition we have that $\mathbb{P}\left(Z(n,t) = t\right) \leq O\left(1\right)\frac{\mathbb{E}[Z(n,t)]}{t}$. We thus obtain

$$\begin{aligned}
&\sum_{t=1}^{N-n} \mathbb{P}\left(Z(n,t) = t\right)\mathbb{P}\left(D_{\mathrm{Po}} = k_n + t\right) \\
&\quad \leq O\left(1\right)\sum_{t=1}^{N-n} \frac{\mathbb{E}\left[Z(n,t)\right]}{t}\left(p_{k_n+t} + n^{-2\alpha(1-\varepsilon)}\right) \\
&\quad = O\left(1\right)\sum_{t=1}^{N-n} \frac{(N-n)(k_n+t)}{Nt}\left(p_{k_n+t} + n^{-2\alpha(1-\varepsilon)}\right) \\
&\quad = O\left(\sqrt{\frac{\log n}{n}}\right)\sum_{t=1}^{N-n} \frac{k_n}{t}\left(p_{k_n+t} + n^{-2\alpha(1-\varepsilon)}\right) \\
&\qquad + O\left(\sqrt{\frac{\log n}{n}}\right)\sum_{t=1}^{N-n}\left(p_{k_n+t} + n^{-2\alpha(1-\varepsilon)}\right),
\end{aligned}$$

where we used that $\frac{N-n}{N} = O\left(\sqrt{\frac{\log n}{n}}\right)$.

We will show that both summations are $o\left(p_{k_n}\right)$. For the first summation we recall that $\frac{k_n(\log n)^{3/2}}{\sqrt{n}} = o\left(1\right)$ for $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$, while for $\varepsilon > 0$ small enough $n^{-2\alpha(1-\varepsilon)} = o\left(k_n^{-(2\alpha+1)}\right) = o\left(p_{k_n}\right)$. Hence, since $p_{k_n+t} \leq p_{k_n}$,

$$\begin{aligned}
O\left(\sqrt{\frac{\log n}{n}}\right)\sum_{t=1}^{N-n} \frac{k_n}{t}\left(p_{k_n+t} + n^{-2\alpha(1-\varepsilon)}\right) &\leq O\left(k_n\sqrt{\frac{\log n}{n}}p_{k_n}\right)\sum_{t=1}^{N-n}\frac{1}{t} \\
&= O\left(\frac{k_n(\log n)^{3/2}}{\sqrt{n}}\right)p_{k_n} = o\left(p_{k_n}\right).
\end{aligned}$$

For the other summation we use that

$$\sum_{t=1}^{N-n} p_{k_n+t} \leq \sum_{t=1}^{\infty} p_{k_n+t} \leq O\left(k_n^{-2\alpha}\right) = O\left(k_n p_{k_n}\right),$$

together with the fact that for ε small enough, $\log(n)n^{-2\alpha(1-\varepsilon)} = o\left(k_n^{-(2\alpha+1)}\right) = o\left(p_{k_n}\right)$. This implies that

$$\begin{aligned}
&O\left(\sqrt{\frac{\log n}{n}}\right)\sum_{t=1}^{N-n}\left(p_{k_n+t}+n^{-2\alpha(1-\varepsilon)}\right)\\
&\le O\left(\sqrt{\frac{\log n}{n}}k_n p_{k_n}\right)+O\left((N-n)\sqrt{\frac{\log n}{n}}n^{-2\alpha(1-\varepsilon)}\right)\\
&\le o\left(p_{k_n}\right)+O\left((\log n)n^{-2\alpha(1-\varepsilon)}\right)=o\left(p_{k_n}\right).
\end{aligned}$$

It now follows that

$$\begin{aligned}
n\mathbb{P}\left(D_n(u)=k_n, D_{\mathrm{Po}}(u)>k_n\right) &= \sum_{t=1}^{N-n}\mathbb{P}\left(D_n=k_n, D_{\mathrm{Po}}=k_n+t\right)\\
&= n\sum_{t=1}^{N-n}\mathbb{P}\left(Z(n,t)=t\right)\mathbb{P}\left(D_{\mathrm{Po}}=k_n+t\right)\\
&= n\, o\left(\mathbb{P}\left(D_{\mathrm{Po}}=k_n\right)\right)=o\left(\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]\right),
\end{aligned}$$

which finishes the proof for the case where $N > n$. ■

3.3.6 First moment for large degrees

Proof of Lemma 3.1.11: Using Stirling's bounds

$$\sqrt{2\pi s}\left(\frac{s}{e}\right)^{-s} \le s! \le e\sqrt{2\pi s}\left(\frac{s}{e}\right)^{-s},$$

we have

$$\begin{aligned}
\mathbb{P}\left(\mathrm{Bin}(n,\lambda/n)=k\right) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{k}\right)^{n-k}\\
&\le \frac{e}{\sqrt{2\pi}}\sqrt{\frac{n}{n-k}}\frac{n^n}{k!}(n-k)^{-(n-k)}e^{-k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\
&= \frac{e}{\sqrt{2\pi}}\sqrt{\frac{n}{n-k}}\frac{\lambda^k e^{-\lambda}}{k!}\left(\frac{n-\lambda}{n-k}\right)^{n-k}e^{\lambda-k}\\
&= \frac{e}{\sqrt{2\pi}}\sqrt{\frac{n}{n-k}}\mathbb{P}\left(\mathrm{Po}(\lambda)=k\right)\left(\frac{n-\lambda}{n-k}\right)^{n-k}e^{\lambda-k}.
\end{aligned}$$

The result then follows by observing that $\left(\frac{n-\lambda}{n-k}\right)^{n-k}e^{\lambda-k} \le 1$ for all $0<\lambda<n$ and $0 \le k \le n-1$. ■

Proof of Lemma 3.1.12: First we observe that as the poissonized KPKVB model G_{Po} has the same intensity measure as the original KPKVB model with a fixed number n of points, the expected degree of a vertex of the KPKVB model with radial coordinate $r = R - y$ is given by $\mu_{Po,n}(y)$ and hence,

$$\mathbb{E}\left[N_n(k_n)\right] = n \int_0^R \mathbb{P}\left(\mathrm{Bin}(n-1, \mu_{Po,n}(y)/n) = k_n\right) \frac{\alpha \sinh(\alpha(R-y))}{\cosh(\alpha R) - 1} \,\mathrm{d}y.$$

Fix $0 < \varepsilon < \frac{4\alpha-1}{4\alpha+2} \wedge \frac{2\alpha-1}{2\alpha}$. We first show that we only need to consider integration up to $y \leq (1-\varepsilon)R$. By our choice of ε, $2\alpha(1-\varepsilon) > 1$, so that

$$\frac{\cosh(\alpha\varepsilon R) - 1}{\cosh(\alpha R) - 1} = O\left(n^{-2\alpha(1-\varepsilon)}\right) = o\left(n^{-1}\right).$$

This implies

$$\begin{aligned} &n \int_{(1-\varepsilon)R}^R \mathbb{P}\left(\mathrm{Bin}(n-1, \mu_{Po,n}(y)/n) = k_n\right) \frac{\alpha \sinh(\alpha(R-y))}{\cosh(\alpha R) - 1} \,\mathrm{d}y \\ &\leq n \frac{\cosh(\alpha \epsilon R)}{\cosh(\alpha R) - 1} = o\left(1\right), \end{aligned}$$

and thus it is enough to show that

$$n \int_0^{(1-\varepsilon)R} \mathbb{P}\left(\mathrm{Bin}(n-1, \mu_{Po,n}(y)) = k_n\right) \frac{\alpha \sinh(\alpha(R-y))}{\cosh(\alpha R) - 1} \,\mathrm{d}y = o\left(1\right).$$

Note that for all $0 \leq y \leq (1-\varepsilon)R$ we have

$$\frac{\alpha \sinh(\alpha(R-y))}{\cosh(\alpha R) - 1} = (1 + o\left(1\right))\alpha e^{-\alpha y}.$$

Hence, by bounding the Binomial probability (see Lemma 3.1.11)

$$\begin{aligned} &n \int_0^{(1-\varepsilon)R} \mathbb{P}\left(\mathrm{Bin}(n-1, \mu_{Po,n}(y)) = k_n\right) \frac{\alpha \sinh(\alpha(R-y))}{\cosh(\alpha R) - 1} \,\mathrm{d}y \\ &\quad \leq (1 + o\left(1\right)) \frac{e}{\sqrt{2\pi}} \sqrt{\frac{n}{n-k_n}} n \int_0^{(1-\varepsilon)R} \mathbb{P}\left(\mathrm{Po}(\mu_{Po,n}(y)) = k_n\right) \alpha e^{-\alpha y} \,\mathrm{d}y \\ &\quad \leq (1 + o\left(1\right)) \frac{e}{\sqrt{2\pi}} \sqrt{\frac{n}{n-k_n}} n p_{k_n} = O\left(\sqrt{\frac{n}{n-k_n}} n k_n^{-(2\alpha+1)}\right). \end{aligned}$$

We shall now consider two cases: $n^{\frac{1}{2\alpha+1}} \ll k_n < n^{1-\varepsilon}$ and $n^{1-\varepsilon} \leq k_n \leq n-1$.

If $n^{\frac{1}{2\alpha+1}} \ll k_n < n^{1-\varepsilon}$ then $\sqrt{\frac{n}{n-k_n}} = 1 + o\left(1\right)$ and hence

$$\sqrt{\frac{n}{n-k_n}} n k_n^{-(2\alpha+1)} = O\left(n k_n^{-(2\alpha+1)}\right) = o\left(1\right).$$

For $k_n \geq n^{1-\varepsilon}$ we have, by our choice of ε, that $\frac{3}{2} - (2\alpha+1)(1-\varepsilon) < 0$, and thus

$$\sqrt{\frac{n}{n-k_n}} n k_n^{-(2\alpha+1)} = O\left(n^{\frac{3}{2}} k_n^{-(2\alpha+1)}\right) = O\left(n^{\frac{3}{2}-(2\alpha+1)(1-\varepsilon)}\right) = o\left(1\right).$$

■

Chapter 4

Local clustering

4.1 Clustering and the degree of the typical point in G_∞

As mentioned earlier, we plan to make use of the Campbell-Mecke formula for comparing the clustering coefficient and function of G_{Po} with certain quantities associated with G_∞. We will be considering the Poisson process $\mathcal{P}$ to which we add one additional point $(0, y)$ on the y-axis. In some computations the height y will be fixed, but eventually we shall take it exponentially distributed with parameter α, and independent of $\mathcal{P}$. We refer to $(0, y)$ as "the typical point".

To provide some intuition for this definition and name, note that we can alternatively view $\mathcal{P}$ as follows. We take a constant intensity Poisson process on $\mathbb{R}$ corresponding to the x-coordinates, and to each point we attach a random "mark", corresponding to the y-coordinate, where the marks are i.i.d. exponentially distributed with parameter α.

Since the clustering coefficient $c(G)$ is defined as an average over all vertices of the graph, it is not immediately obvious how to meaningfully define a corresponding notion for infinite graphs, and similarly for the clustering function. We can however without any issues speak of the (expected) clustering coefficient of the typical point, or the expected clustering coefficient given that it has degree k, or the distribution of the degree of the typical point.

If $p = (x, y) \in \mathbb{R} \times [0, \infty)$ is a point, not necessarily part of the Poisson process, recall that $\mathcal{B}_\infty(p)$ denotes its neighbourhood ball in the infinite limit model (see list of notation). For such a point, we will write

$$\mu(y) = \mu(p) := \mu(\mathcal{B}_\infty(p)).$$

Integrating the intensity function of $\mathcal{P}$ over $\mathcal{B}_\infty(p)$ gives, using $\alpha > \frac{1}{2}$,

$$\mu(y) = \int_{\mathcal{B}_\infty(p)} f(x', y')\, dx'\, dy' = \int_0^\infty \int_{-e^{(y+y')/2}}^{e^{(y+y')/2}} \frac{\alpha\nu}{\pi} e^{-\alpha y'}\, dx'\, dy'$$

$$= \int_0^\infty 2e^{(y+y')/2}\frac{\alpha\nu}{\pi}e^{-\alpha y'}\,dy' = \frac{2\alpha\nu e^{y/2}}{\pi}\int_0^\infty e^{(\frac{1}{2}-\alpha)y'}\,dy'$$
$$= \frac{2\alpha\nu e^{y/2}}{\pi(\alpha-\frac{1}{2})} = \xi e^{y/2}.$$

4.1.1 The degree of the typical point

Before considering clustering we briefly investigate the distribution of the degree of the typical point. For $p = (x, y) \in \mathbb{R} \times [0, \infty)$ we define

$$\rho(p,k) := \mathbb{P}\left(\mathrm{Po}(\mu\left(p\right)) = k\right), \tag{4.1}$$

where $\mathrm{Po}(\lambda)$ denotes a Poisson random variable with expectation λ. We will often write $\rho(y, k)$ instead of $\rho(p, k)$.

Let the random variable D denote the degree of the typical point. Since the typical point has a height that is independent of the Poisson process and $\mathrm{Exp}(\alpha)$-distributed, for $k \in \mathbb{N}_0$:

$$p_k := \mathbb{P}\left(D = k\right) = \int_0^\infty \rho(y,k)\alpha e^{-\alpha y}\,dy. \tag{4.2}$$

Using the transformation of variables $z = \xi e^{\frac{y}{2}}$ (so $dy = \frac{2}{z}dz$), we compute

$$\begin{aligned} p_k &= \frac{1}{k!}\int_0^\infty \left(\xi e^{\frac{y}{2}}\right)^k e^{-\xi e^{\frac{y}{2}}}\alpha e^{-\alpha y}\,dy = \frac{\alpha\xi^{2\alpha}}{k!}\int_0^\infty \left(\xi e^{\frac{y}{2}}\right)^{k-2\alpha} e^{-\xi e^{\frac{y}{2}}}\,dy \\ &= \frac{2\alpha\xi^{2\alpha}}{k!}\int_\xi^\infty z^{k-2\alpha-1}e^{-z}\,dz = \frac{2\alpha\xi^{2\alpha}\Gamma^+(k-2\alpha,\xi)}{k!}, \end{aligned} \tag{4.3}$$

where we recall that Γ denotes the gamma function and Γ^+ the upper incomplete gamma function. Note that, unsurprisingly, this is identical to the expression Gugelmann et al. [29] gave for the limiting degree distribution of $G(n; \alpha, \nu)$. Using Stirling's approximation to the gamma function, we find that

$$p_k \sim 2\alpha\xi^{2\alpha}k^{-(2\alpha+1)} \quad \text{as } k \to \infty. \tag{4.4}$$

By a similar computation we have the following result, which will be useful later on. For any $\beta > 0$, as $k \to \infty$,

$$\int_0^\infty e^{-\beta y}\rho(y,k)\alpha e^{-\alpha y}\,dy \sim 2\alpha\xi^{2(\beta+\alpha)}k^{-2(\beta+\alpha)-1}. \tag{4.5}$$

4.1.2 The expected clustering coefficient and function of the typical point

Let the random variable C denote the clustering coefficient of the typical point $(0, y)$, in the graph obtained from G_∞ by adding $(0, y)$. We now *define*

$$\gamma := \mathbb{E}[C],$$

(where we take the expectation over both the Poisson point process $\mathcal{P}$ and $y \stackrel{\mathrm{d}}{=} \mathrm{Exp}(\alpha)$, independently of the Poisson process $\mathcal{P}$). We shall show shortly that γ takes on the value stated in Theorem 1.5.1.

For any fixed value $y_0 > 0$, the set of points inside $\mathcal{B}_\infty((0, y_0))$ is a Poisson process with intensity $f \cdot \mathbb{1}_{\mathcal{B}_\infty((0,y_0))}$. As $\mu(\mathcal{B}_\infty((0, y_0))) = \mu(y_0) = \xi e^{y_0/2} < \infty$, this can be described alternatively by first picking $N \stackrel{\mathrm{d}}{=} \mathrm{Po}(\mu(y_0))$ and then taking N i.i.d. points in $\mathcal{B}_\infty((0, y_0))$ according to the probability density $f \cdot \mathbb{1}_{\mathcal{B}_\infty((0,y_0))}/\mu(y_0)$. (That is, the intensity function of the Poisson point process, but set to zero outside of $\mathcal{B}_\infty((0, y_0))$ and re-normalized in such a way that it integrates to one.) Hence, if we condition on the event that y takes on some fixed value y_0 and that there are exactly k points of $\mathcal{P}$ inside $\mathcal{B}_\infty((0, y_0))$, then those k points behave like k i.i.d. points in $\mathcal{B}_\infty((0, y_0))$ chosen according to the mentioned re-normalized probability density function. This shows that, for every $k \geq 2$:

$$\mathbb{E}\left[C \mid D = k, y = y_0\right] = \frac{1}{\binom{k}{2}} \mathbb{E}\left[\sum_{1 \leq i < j \leq k} \mathbb{1}_{\{u_i \in \mathcal{B}_\infty(u_j)\}}\right] = \mathbb{E}\left[\mathbb{1}_{\{u_1 \in \mathcal{B}_\infty(u_2)\}}\right],$$

where $u_1, \ldots, u_k$ are i.i.d. points in $\mathcal{B}_\infty((0, y_0))$ with the above mentioned density. Note that this does not depend on the value of k. For notational convenience, we will write

$$P(y_0) := \mathbb{E}\left[\mathbb{1}_{\{u_1 \in \mathcal{B}_\infty(u_2)\}}\right] = \mathbb{E}\left[C \mid D = k, y = y_0\right],$$

with u_1, u_2 as above.

We now observe that

$$\begin{aligned}\gamma(k) := \mathbb{E}\left[C \mid D = k\right] &= \int_0^\infty \mathbb{E}\left[C \mid D = k, y = y_0\right] h_k(y_0)\, dy_0 \\ &= \int_0^\infty P(y_0) h_k(y_0)\, dy_0,\end{aligned}$$

where h_k denotes the density of y *conditional on* $D = k$. That is,

$$h_k(y_0) = \frac{\rho(y_0, k)\alpha e^{-\alpha y_0}}{\int_0^\infty \rho(t, k)\alpha e^{-\alpha t}\, dt} = \frac{1}{p_k} \cdot \rho(y_0, k)\alpha e^{-\alpha y_0}.$$

Hence,

$$\gamma(k) = \frac{1}{p_k} \cdot \int_0^\infty P(y_0)\rho(y_0, k)\alpha e^{-\alpha y_0}\, dy_0. \tag{4.6}$$

This also gives

$$\begin{aligned}\gamma &= \mathbb{E}\left[C\right] = \sum_{k \geq 2} \mathbb{E}\left[C \mid D = k\right] \mathbb{P}\left(D = k\right) \\ &= \int_0^\infty P(y_0) \left(\sum_{k=2}^\infty \rho(y_0, k)\right) \alpha e^{-\alpha y_0}\, dy_0 \qquad (4.7) \\ &= \int_0^\infty P(y_0) \left(1 - \rho(y_0, 0) - \rho(y_0, 1)\right) \alpha e^{-\alpha y_0}\, dy_0.\end{aligned}$$

A key step is to derive the following explicit expression for $P(y_0)$:

Lemma 4.1.1. *If $\alpha \neq 1$, then*

$$
\begin{aligned}
P(y_0) = &-\frac{1}{8(\alpha-1)\alpha} + \frac{(\alpha-1/2)e^{-\frac{1}{2}y_0}}{\alpha-1} - \frac{(\alpha-1/2)^2 e^{-y_0}}{4(\alpha-1)^2} \\
&+ (e^{-\frac{1}{2}y_0})^{4\alpha-2}\left(\frac{2^{-4\alpha-1}(3\alpha-1)}{\alpha(\alpha-1)^2} + \frac{(\alpha-1/2)B^-(1/2; 1+2\alpha, -2+2\alpha)}{2(\alpha-1)\alpha}\right) \\
&+ \frac{(1-e^{-\frac{1}{2}y_0})^{2\alpha}}{8(\alpha-1)\alpha} - \frac{(e^{-\frac{1}{2}y_0})^{4\alpha-2}B^-(1-e^{-\frac{1}{2}y_0}; 2\alpha, 3-4\alpha)}{4(\alpha-1)},
\end{aligned}
$$

where we recall that B^- denotes the lower incomplete beta function.

We will prove this lemma in a sequence of steps.

Recall that $P(y_0)$ is the probability that $u_1 = (x_1, y_1), u_2 = (x_2, y_2)$ are neighbours in G_∞, where u_1, u_2 are i.i.d. with probability density $f \cdot \mathbb{1}_{\mathcal{B}_\infty((0,y_0))}/\mu(y_0)$. In particular,

$$
\begin{aligned}
\mathbb{P}(y_i > t) &= \frac{\nu\alpha}{\pi\mu(y_0)} \int_t^\infty \int_{-e^{(y+y_0)/2}}^{e^{(y+y_0)/2}} e^{-\alpha y}\, \mathrm{d}x\, \mathrm{d}y = \frac{\nu\alpha}{\pi\mu(y_0)} \int_t^\infty 2e^{(y+y_0)/2} \cdot e^{-\alpha y}\, \mathrm{d}y \\
&= \frac{2\nu\alpha e^{y_0/2}}{\pi\xi e^{y_0/2}(\alpha-\frac{1}{2})} \cdot e^{(\frac{1}{2}-\alpha)t} = e^{(\frac{1}{2}-\alpha)t},
\end{aligned}
$$

using that $\mu(y_0) = \xi e^{y_0/2} = \left(\frac{2\alpha\nu}{\pi(\alpha-\frac{1}{2})}\right) e^{y_0/2}$. Thus, y_1, y_2 are exponentially distributed with parameter $\alpha - \frac{1}{2} > 0$. Now note that, for each $t > 0$, the probability density $f \cdot \mathbb{1}_{\mathcal{B}_\infty((0,y_0))}/\mu(y_0)$ is constant on $[-e^{(t+y_0)/2}, e^{(t+y_0)/2}] \times \{t\}$ and it vanishes on $(-\infty, -e^{(t+y_0)/2}) \times \{t\} \cup (e^{(t+y_0)/2}, \infty) \times \{t\}$.

Hence, given the height y_i of u_i, the x-coordinate of u_i is uniformly distributed in $[-e^{\frac{1}{2}(y+y_i)}, e^{\frac{1}{2}(y+y_i)}]$. With this in mind we define $P(y_0, y_1, y_2)$ to be the probability that $(0, y_0), (x_1, y_1), (x_2, y_2)$ satisfy $|x_1 - x_2| \leq e^{(y_1+y_2)/2}$, where x_1 and x_2 are independent uniform random variables in the intervals $[-e^{\frac{1}{2}(y_0+y_1)}, e^{\frac{1}{2}(y_0+y_1)}]$ and $[-e^{\frac{1}{2}(y_0+y_2)}, e^{\frac{1}{2}(y_0+y_2)}]$, respectively. Then, we have that

$$
P(y_0) = (\alpha-1/2)^2 \int_0^\infty \int_0^\infty P(y_0, y_1, y_2) e^{-(\alpha-1/2)(y_1+y_2)}\, \mathrm{d}y_2\, \mathrm{d}y_1. \tag{4.8}
$$

Determining $P(y_0, y_1, y_2)$

To compute the integral (4.8) it will be convenient to use the change of variable $z_i = e^{-y_i/2}$, for $i = 0, 1, 2$. We will write $y_i(z_i)$ to stress the dependence between y_i and z_i. The following result completely characterizes $P(y_0, y_1, y_2)$:

Lemma 4.1.2.

$$P(y_0(z_0), y_1(z_1), y_2(z_2)) = \begin{cases} 1, & \text{if } z_0 \geq z_1 + z_2, z_0 > z_1 > z_2, \\ 1 - G(z_0, z_1, z_2), & \text{if } z_0 < z_1 + z_2, z_0 > z_1 > z_2, \\ \frac{z_0}{z_1}, & \text{if } z_1 \geq z_0 + z_2, z_1 > \max(z_0, z_2), \\ \frac{z_0}{z_1}\left(1 - G(z_1, z_0, z_2)\right), & \text{if } z_1 < z_0 + z_2, z_1 > \max(z_0, z_2), \end{cases}$$

where

$$G(a,b,c) = \frac{1}{4}\left(b^{-1}c + bc^{-1} + a^2b^{-1}c^{-1} + 2 - 2ab^{-1} - 2ac^{-1}\right).$$

We split the proof of this lemma into a couple of smaller pieces. We begin with the following lemma.

Lemma 4.1.3. *Let $z_i = e^{-y_i/2}$, $i = 0, 1, 2$. If $y_0 < y_1 < y_2$ (or equivalently $z_0 > z_1 > z_2$), then*

$$P(y_0(z_0), y_1(z_1), y_2(z_2)) = \begin{cases} 1, & \text{if } z_0 \geq z_1 + z_2, \\ 1 - G(z_0, z_1, z_2), & \text{if } z_0 < z_1 + z_2. \end{cases}$$

Proof. Note that $P(y_0, y_1, y_2)$ is the probability that x_2 falls into the interval $[x_1 - e^{(y_1+y_2)/2}, x_1 + e^{(y_1+y_2)/2}]$, as well as into the interval $[-e^{(y_0+y_2)/2}, e^{(y_0+y_2)/2}]$. By symmetry considerations, we can take x_1 uniformly at random from $[0, e^{y_0/2+y_1/2}]$ as opposed to $[-e^{y_0/2+y_1/2}, e^{y_0/2+y_1/2}]$. Figure 4.1 shows the intersection of the intervals (red line) for two different cases for $x_1 \leq e^{(y_0+y_1)/2}$.

Since $y_0 < y_1 < y_2$ we have that $e^{(y_1+y_2)/2} > e^{(y_0+y_2)/2}$ and so, when $x_1 \geq 0$, the "right half" of the interval $[-e^{(y_0+y_2)/2}, e^{(y_0+y_2)/2}]$ is always covered by the interval $[x_1 - e^{(y_1+y_2)/2}, x_1 + e^{(y_1+y_2)/2}]$. If $e^{(y_1+y_2)/2} - e^{(y_0+y_1)/2} \geq e^{(y_0+y_2)/2}$ then the "left half" is always covered as well. In other words:

$$e^{(y_1+y_2)/2} - e^{(y_0+y_1)/2} \geq e^{(y_0+y_2)/2} \Rightarrow P(y_0, y_1, y_2) = 1.$$

Now consider the case where $e^{(y_1+y_2)/2} - e^{(y_0+y_1)/2} < e^{(y_0+y_2)/2}$. Then, if $x_1 \in [0, e^{(y_1+y_2)/2} - e^{(y_0+y_2)/2}]$ the whole interval $[-e^{(y_0+y_2)/2}, e^{(y_0+y_2)/2}]$ is still covered so that p_0, p_1 and p_2 form a triangle. If, on the other hand $e^{(y_1+y_2)/2} - e^{(y_0+y_2)/2} < x_1 \leq e^{(y_0+y_1)/2}$ then the probability that $|x_2 - x_1| \leq e^{(y_1+y_2)/2}$ equals

$$1 - \frac{x_1 - (e^{(y_1+y_2)/2} - e^{(y_0+y_2)/2})}{2e^{(y_0+y_2)/2}}.$$

Hence, if $e^{(y_1+y_2)/2} - e^{(y_0+y_1)/2} < e^{(y_0+y_2)/2}$ we have

$P(y_0, y_1, y_2)$

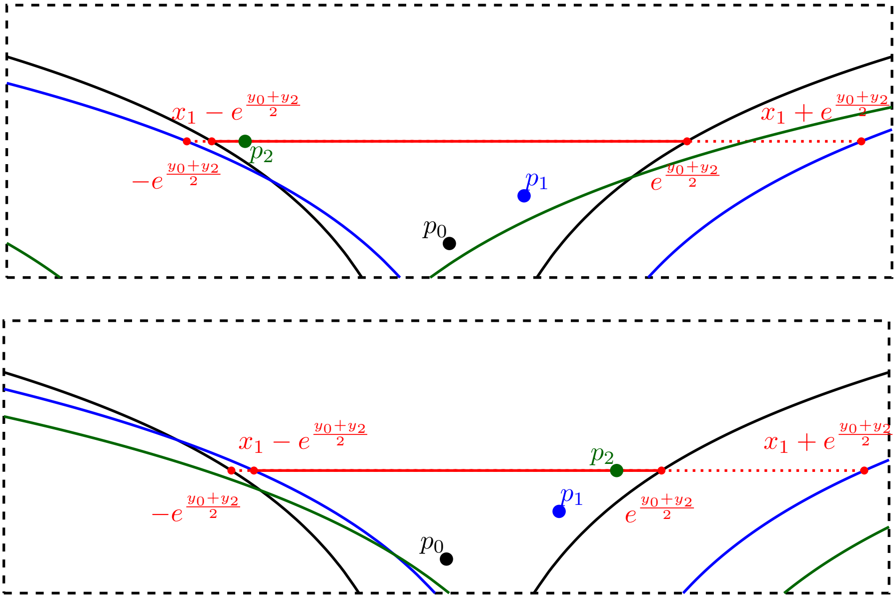

Figure 4.1: Situation for the intersections of the connection intervals considered in Lemma 4.1.3, with $y_0 < y_1 < y_2$ fixed and for different cases of $0 \leq x_1 \leq e^{(y_0+y_1)/2}$. The top figure shows the case where $0 \leq x_1 \leq e^{(y_1+y_2)/2} - e^{(y_0+y_2)/2}$, while the bottom one shows the case $x_1 > e^{(y_1+y_2)/2} - e^{(y_0+y_2)/2}$. The solid red line indicates the range for x_2 such that the points p_0, p_1 and p_2 form a triangle. The boundaries of their neighbourhoods are shown in, respectively, black, blue and green.

$$
\begin{aligned}
&= \frac{e^{(y_1+y_2)/2} - e^{(y_0+y_2)/2}}{e^{(y_0+y_1)/2}} \\
&\quad + \int_{e^{(y_1+y_2)/2}-e^{(y_0+y_2)/2}}^{e^{(y_0+y_1)/2}} \left(1 - \frac{x_1 - (e^{(y_1+y_2)/2} - e^{(y_0+y_2)/2)})}{2e^{(y_0+y_2)/2}}\right) \cdot \frac{1}{e^{(y_0+y_1)/2}} \,\mathrm{d}x_1 \\
&= 1 - \frac{1}{2e^{y_0+y_1/2+y_2/2}} \int_0^{e^{(y_0+y_1)/2}+e^{(y_0+y_2)/2}-e^{(y_1+y_2)/2}} x_1 \,\mathrm{d}x_1 \\
&= 1 - \frac{\left(e^{(y_0+y_1)/2} + e^{(y_0+y_2)/2} - e^{(y_1+y_2)/2}\right)^2}{4e^{y_0+y_1/2+y_2/2}}.
\end{aligned}
$$

At this point, it is convenient to rewrite everything in terms of $z_i := e^{-y_i/2}$. Note that $y_0 < y_1 < y_2$ if and only if $z_0 > z_1 > z_2$ while the condition $e^{(y_1+y_2)/2} - e^{(y_0+y_1)/2} < e^{(y_0+y_2)/2}$ becomes

$$
e^{(y_1+y_2)/2} - e^{(y_0+y_1)/2} < e^{(y_0+y_2)/2} \Leftrightarrow z_1^{-1} z_2^{-1} < z_0^{-1} z_1^{-1} + z_0^{-1} z_2^{-1} \Leftrightarrow z_0 < z_1 + z_2.
$$

We now conclude that

$$P(y_0(z_0), y_1(z_1), y_2(z_2)) = 1 \quad \text{if} \quad z_0 > z_1 > z_2 \text{ and } z_0 \geq z_1 + z_2$$

while for $z_0 > z_1 > z_2$ and $z_0 < z_1 + z_2$

$$\begin{aligned} &P(y_0(z_0), y_1(z_1), y_2(z_2)) \\ &\quad = 1 - \frac{z_0^2 z_1 z_2}{4} \cdot \left(z_0^{-1} z_1^{-1} + z_0^{-1} z_2^{-1} - z_1^{-1} z_2^{-1}\right)^2 \\ &\quad = 1 - \frac{1}{4}\left(z_1^{-1} z_2 + z_1 z_2^{-1} + z_0^2 z_1^{-1} z_2^{-1} + 2 - 2z_0 z_1^{-1} - 2z_0 z_2^{-1}\right), \end{aligned}$$

which finishes the proof. □

The previous lemma covers the case when $y_0 < y_1 < y_2$. We now leverage it to take care of the other cases as well.

Proof of Lemma 4.1.2. Let $y_i > 0$ and $z_i = e^{-y_i/2}$, $i = 0, 1, 2$. Lemma 4.1.3 gives the expression for $P(y_0(z_0), y_1(z_1), y_2(z_2))$ in the case $y_0 < y_1 < y_2$, or equivalently $z_0 > z_1 > z_2$, i.e. the first two lines in the claim of Lemma 4.1.2. To analyze the other cases we shall express $P(y_1, y_0, y_2)$ and $P(y_1, y_2, y_0)$ in terms of $P(y_0, y_1, y_2)$ and z_i. For this we note that we can view $P(y_0, y_1, y_2)$ as a 2-fold integral of the indicator function

$$h(x_0, x_1, x_2) := \mathbb{1}_{\left\{|x_0 - x_1| < e^{(y_0+y_1)/2}, |x_0 - x_2| < e^{(y_0+y_2)/2}, |x_1 - x_2| < e^{(y_1+y_2)/2}\right\}},$$

where x_0 was set to zero, without loss of generality, and the other two x_i are uniform random variables on $[-e^{(y_0+y_i)/2}, e^{(y_0+y_i)/2}]$. When we consider the probability $P(y_1, y_0, y_2)$, this is the 2-fold integral of $h(x_0, 0, x_2)$ so that

$$\begin{aligned} P(y_1, y_0, y_2) &= \frac{1}{2e^{(y_1+y_0)/2}} \cdot \frac{1}{2e^{(y_1+y_2)/2}} \iint_{\mathbb{R}} h(x_0, 0, x_2) \, \mathrm{d}x_0 \, \mathrm{d}x_2 \\ &= \frac{e^{y_0/2}}{e^{y_1/2}} \frac{1}{2e^{(y_0+y_1)/2}} \frac{1}{2e^{(y_0+y_2)/2}} \iint_{\mathbb{R}} h(0, x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \frac{e^{y_0/2}}{e^{y_1/2}} P(y_0, y_1, y_2) = \frac{z_1}{z_0} P(y_0, y_1, y_2). \end{aligned}$$

Finally we note that $h(x_0, 0, x_2) = h(x_2, 0, x_0)$ from which we conclude that

$$P(y_0, y_1, y_2) = (z_0/z_1)\, P(y_1, y_0, y_2) = (z_0/z_1)\, P(y_1, y_2, y_0). \tag{4.9}$$

To complete the proof for the other cases we note that since $P(y_0, y_1, y_2)$ is symmetric in y_1 and y_2, we can assume, without loss of generality, that $y_1 < y_2$. Then, there are two more orderings of y_0, y_1, y_2, namely $y_1 < y_0 < y_2$ and $y_1 < y_2 < y_0$, which can be summarized as $y_1 < \min(y_0, y_2)$, or equivalently $z_1 > \max(z_0, z_2)$. For $y_1 < y_0 < y_2$ and $y_1 < y_2 < y_0$ we can apply Lemma 4.1.3 to obtain $P(y_1, y_0, y_2) = P(y_1, y_2, y_0)$ which happens to agree due to the symmetry in the last two arguments of the expression found in Lemma 4.1.3. The expression for $P(y_0, y_1, y_2)$ then follows from (4.9). □

Integrating over y_1, y_2

Now that we have established the expression for $P(y_0, y_1, y_2)$ we can proceed to compute $P(y_0)$ by integrating over y_1, y_2. We however start with the following observation.

Lemma 4.1.4. *The function $\alpha \mapsto P_\alpha(y_0)$ is continuous for all $\alpha > \frac{1}{2}$.*

Proof. This follows from the theorem of dominated convergence: Let $\alpha > \frac{1}{2}$ and $(\alpha_n)_{n\in\mathbb{N}}$ a sequence of real numbers converging to α, so we can assume $|\alpha_n - \alpha| < \epsilon := \frac{\alpha-1/2}{2}$. This means that $-\epsilon < \alpha_n - \alpha < \epsilon$, i.e. $\frac{\alpha-1/2}{2} < \alpha_n - 1/2 < \frac{3\alpha-3/2}{2}$. Define

$$f_n(y_1, y_2) = P(y_0, y_1, y_2)(\alpha_n - 1/2)^2 e^{-(\alpha_n-1/2)(y_1+y_2)}.$$

As the function $x \mapsto x^2$ is increasing in x for $x > 0$ and the function $x \mapsto e^{-(y_1+y_2)x}$ is decreasing in x and $P(y_0, y_1, y_2) \in [0, 1]$, it holds that

$$|f_n(y_1, y_2)| \leq \left(\frac{3\alpha - 3/2}{2}\right)^2 e^{-(y_1+y_2)\frac{\alpha-1/2}{2}},$$

which is integrable over $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ (with integral equalling $(6\alpha-3)^2/(2\alpha-1)^2$). Application of the theorem of dominated convergence yields that $P_{\alpha_n}(y_0) \to P_\alpha(y_0)$ which gives the claim as the sequence $(\alpha_n)_n$ was arbitrary. $\square$

Due to this lemma we can first assume $\alpha \notin \{\frac{3}{4}, 1\}$, compute $P(y_0)$ and then obtain the values of $P(y_0)$ at the remaining two points by taking the corresponding limit in α. This strategy is executed below. It involves the computation of several integrals which are involved and will take up a few pages. The proof is structured using headers, to aid the reader.

Note that when writing $P(y_0)$ as an integral, see equation (4.8), by symmetry in the integration variables y_1 and y_2, we can assume that $y_1 < y_2$ in which case either y_0 or y_1 is the smallest height. This gives half the value of $P(y_0)$ and hence

$$P(y_0) = 2(I_1(y_0) + I_2(y_0)),$$

where I_1 and I_2 are given by

$$I_1(y_0) := \int_{0<y_0<y_1<y_2} P(y_0, y_1, y_2) \cdot (\alpha - 1/2)^2 e^{-(\alpha-1/2)(y_1+y_2)} \, \mathrm{d}y_2 \, \mathrm{d}y_1,$$
$$I_2(y_0) := \int_{0<y_1<y_0,y_2} P(y_0, y_1, y_2) \cdot (\alpha - 1/2)^2 e^{-(\alpha-1/2)(y_1+y_2)} \, \mathrm{d}y_2 \, \mathrm{d}y_1.$$

We proceed with computing these two integrals, each of which is split into two parts. The final expressions of those four integrals can be found in (4.10), (4.15), (4.16) and (4.18).

Computing $I_1(y_0)$. Applying the change of variables $z_i := e^{-y_i/2}$, $i = 1, 2$, and Lemma 4.1.2 gives

$$\begin{aligned}
I_1(y_0) &= 4(\alpha - 1/2)^2 \cdot \int_{z_0>z_1>z_2>0} P(y_0, y_1(z), y_2(z)) z_1^{2\alpha-2} z_2^{2\alpha-2}\,\mathrm{d}z_2\,\mathrm{d}z_1 \\
&= 4(\alpha - 1/2)^2 \cdot \Bigg(\int_{z_0>z_1>z_2>0} 1 \cdot z_1^{2\alpha-2} z_2^{2\alpha-2}\,\mathrm{d}z_2\,\mathrm{d}z_1 \\
&\quad - \int_{\substack{z_0>z_1>z_2>0,\\ z_0<z_1+z_2}} G(z_0, z_1, z_2) \cdot z_1^{2\alpha-2} z_2^{2\alpha-2}\,\mathrm{d}z_2\,\mathrm{d}z_1 \Bigg) \\
&=: 4(\alpha - 1/2)^2 (I_{1,1}(y_0) - I_{1,2}(y_0)).
\end{aligned}$$

The integral $I_{1,1}(y_0)$ is easily obtained:

$$\begin{aligned}
I_{1,1}(y_0) &= \int_0^{z_0} \int_0^{z_1} z_1^{2\alpha-2} z_2^{2\alpha-2}\,\mathrm{d}z_2\,\mathrm{d}z_1 = \int_0^{z_0} z_1^{2\alpha-2} \left[\frac{z_2^{2\alpha-1}}{2\alpha - 1} \right]_0^{z_1} \mathrm{d}z_1 \\
&= \frac{1}{2\alpha - 1} \cdot \int_0^{z_0} z_1^{4\alpha-3}\,\mathrm{d}z_1 = \frac{1}{2(2\alpha-1)^2} \cdot z_0^{4\alpha-2}.
\end{aligned} \tag{4.10}$$

To deal with $I_{1,2}$ we note that $G(z_0, z_1, z_2)$ is a linear combination of monomials of the form $z_0^a z_1^b z_2^c$ with $a, b, c \in \{-1, 0, 1, 2\}$ and $a + b + c = 0$. Let us consider the integral $J_{(a,b,c)}(z_0)$ defined by

$$J_{a,b,c}(z_0) := z_0^a \int_{\substack{z_0>z_1>z_2>0,\\ z_0<z_1+z_2}} z_1^{b+2\alpha-2} z_2^{c+2\alpha-2}\,\mathrm{d}z_2\,\mathrm{d}z_1. \tag{4.11}$$

and note that

$$\begin{aligned}
I_{1,2}(y_0) = \frac{1}{4}(&J_{0,-1,1}(z_0) + J_{0,1,-1}(z_0) + J_{2,-1,-1}(z_0) \\
&+ 2J_{0,0,0}(z_0) - 2J_{1,-1,0}(z_0) - 2J_{1,0,-1}(z_0)).
\end{aligned} \tag{4.12}$$

Next we compute $J_{a,b,c}(z_0)$.

$$\begin{aligned}
J_{a,b,c}(z_0) &= z_0^a \int_{z_0/2}^{z_0} \int_{z_0-z_1}^{z_1} z_1^{b+2\alpha-2} z_2^{c+2\alpha-2}\,\mathrm{d}z_2\,\mathrm{d}z_1 \\
&= z_0^a \int_{z_0/2}^{z_0} z_1^{b+2\alpha-2} \left[\frac{z_2^{c+2\alpha-1}}{c + 2\alpha - 1} \right]_{z_0-z_1}^{z_1} \mathrm{d}z_1 \\
&= \frac{z_0^a}{c + 2\alpha - 1} \cdot \left(\int_{z_0/2}^{z_0} z_1^{b+c+4\alpha-3}\,\mathrm{d}z_1 - \int_{z_0/2}^{z_0} z_1^{b+2\alpha-2} (z_0 - z_1)^{c+2\alpha-1}\,\mathrm{d}z_1 \right) \\
&= \frac{z_0^{a+b+c+4\alpha-2}(1 - (1/2)^{b+c+4\alpha-2})}{(c + 2\alpha - 1)(b + c + 4\alpha - 2)} \\
&\quad - \frac{z_0^{a+b+c+4\alpha-3}}{c + 2\alpha - 1} \int_{z_0/2}^{z_0} (z_1/z_0)^{b+2\alpha-2} (1 - (z_1/z_0))^{c+2\alpha-1}\,\mathrm{d}z_1
\end{aligned}$$

$$
\begin{aligned}
&= \frac{z_0^{4\alpha-2}(1-(1/2)^{b+c+4\alpha-2})}{(c+2\alpha-1)(b+c+4\alpha-2)} - \frac{z_0^{4\alpha-2}}{c+2\alpha-1}\int_{1/2}^{1} u^{b+2\alpha-2}(1-u)^{c+2\alpha-1}\,\mathrm{d}u \\
&= \frac{z_0^{4\alpha-2}(1-(1/2)^{b+c+4\alpha-2})}{(c+2\alpha-1)(b+c+4\alpha-2)} - \frac{z_0^{4\alpha-2}}{c+2\alpha-1}B^-(1/2; c+2\alpha, b+2\alpha-1),
\end{aligned}
$$

where we have used the substitution $u := z_1/z_0$ giving $z_0\,\mathrm{d}u = \mathrm{d}z_1$ in the penultimate line and B^- denotes the (lower) incomplete beta function. Note that since $c \geq -1$, $-a \in \{0, -1, -2\}$ and by our assumption $\alpha \notin \{\frac{3}{4}, 1\}$, the denominators that occur during the integration are all non-zero.

Plugging this back into (4.12) gives

$$
\begin{aligned}
&I_{1,2}(y_0) \\
&= \frac{z_0^{4\alpha-2}(1-(1/2)^{4\alpha-2})}{32\alpha(\alpha-1/2)} - \frac{z_0^{4\alpha-2}}{8\alpha}B^-(1/2; 1+2\alpha, 2\alpha-2) \\
&\quad + \frac{z_0^{4\alpha-2}(1-(1/2)^{4\alpha-2})}{32(\alpha-1)(\alpha-1/2)} - \frac{z_0^{4\alpha-2}}{4(2\alpha-2)}B^-(1/2; 2\alpha-1, 2\alpha) \\
&\quad + \frac{z_0^{4\alpha-2}(1-(1/2)^{4\alpha-4})}{32(\alpha-1)^2} - \frac{z_0^{4\alpha-2}}{4(2\alpha-2)}B^-(1/2; -1+2\alpha, 2\alpha-2) \\
&\quad + \frac{z_0^{4\alpha-2}(1-(1/2)^{4\alpha-2})}{16(\alpha-1/2)^2} - \frac{z_0^{4\alpha-2}}{2(2\alpha-1)}B^-(1/2; 2\alpha, 2\alpha-1) \\
&\quad - \frac{z_0^{4\alpha-2}(1-(1/2)^{4\alpha-3})}{16(\alpha-1/2)(\alpha-3/4)} + \frac{z_0^{4\alpha-2}}{2(2\alpha-1)}B^-(1/2; 2\alpha, 2\alpha-2) \\
&\quad - \frac{z_0^{4\alpha-2}(1-(1/2)^{4\alpha-3})}{16(\alpha-1)(\alpha-3/4)} + \frac{z_0^{4\alpha-2}}{2(2\alpha-2)}B^-(1/2; -1+2\alpha, 2\alpha-1) \\
&= \frac{\left(\frac{3}{64} - \frac{3}{16}2^{-4\alpha} + \alpha(-\frac{41}{128} + \frac{13}{16}2^{-4\alpha}) + \alpha^2(\frac{5}{8} - \frac{3}{4}2^{-4\alpha}) - \frac{15}{32}\alpha^3 + \frac{1}{8}\alpha^4\right) z_0^{4\alpha-2}}{4(\alpha-1/2)^2(\alpha-1)^2(\alpha-3/4)\alpha} \\
&\quad + \frac{z_0^{4\alpha-2}}{8(\alpha-1)\alpha(2\alpha-1)}(4(\alpha-1)\alpha(B^-(1/2; 2\alpha, 2\alpha-2) - B^-(1/2; 2\alpha, 2\alpha-1)) \\
&\quad - (2\alpha-1)\alpha(B^-(1/2; 2\alpha-1, 2\alpha-2) + B^-(1/2; 2\alpha-1, 2\alpha) \\
&\qquad\qquad - 2B^-(1/2; 2\alpha-1, 2\alpha-1)) \\
&\quad - (2\alpha-1)(\alpha-1)B^-(1/2; 1+2\alpha, 2\alpha-2)) \\
&= \frac{\left(\frac{3}{64} - \frac{3}{16}2^{-4\alpha} + \alpha(-\frac{41}{128} + \frac{13}{16}2^{-4\alpha}) + \alpha^2(\frac{5}{8} - \frac{3}{4}2^{-4\alpha}) - \frac{15}{32}\alpha^3 + \frac{1}{8}\alpha^4\right) z_0^{4\alpha-2}}{4(\alpha-1/2)^2(\alpha-1)^2(\alpha-3/4)\alpha} \\
&\quad + \frac{z_0^{4\alpha-2}}{8(\alpha-1)\alpha(2\alpha-1)}(4(\alpha-1)\alpha B^-(1/2; 2\alpha+1, 2\alpha-2) \\
&\quad - (2\alpha-1)\alpha B^-(1/2; 2\alpha+1, 2\alpha-2) \\
&\quad - (2\alpha-1)(\alpha-1)B^-(1/2; 2\alpha+1, 2\alpha-2)).
\end{aligned}
$$

For the last step we use the identities

$$B^-(z;a,b) - B^-(z;a,b+1) = B^-(z;a+1,b), \tag{4.13}$$

$$B^-(z;a,b) + B^-(z;a,b+2) - 2B^-(z;a,b+1) = B^-(z;a+2,b). \tag{4.14}$$

to obtain

$$\begin{aligned}
&I_{1,2}(y_0)\\
&= \frac{\left(\frac{3}{64} - \frac{3}{16}2^{-4\alpha} + \alpha(-\frac{41}{128} + \frac{13}{16}2^{-4\alpha}) + \alpha^2(\frac{5}{8} - \frac{3}{4}2^{-4\alpha}) - \frac{15}{32}\alpha^3 + \frac{1}{8}\alpha^4\right) z_0^{4\alpha-2}}{4(\alpha-1/2)^2(\alpha-1)^2(\alpha-3/4)\alpha}\\
&\quad - \frac{z_0^{4\alpha-2}B^-(1/2;2\alpha+1,2\alpha-2)}{8(\alpha-1)\alpha(2\alpha-1)}.
\end{aligned} \tag{4.15}$$

Computing $I_2(y_0)$. We will follow a similar strategy as for $I_1(y_0)$. First, using the change of variables $z_i := e^{-y_i/2}$, $i = 1,2$, we get

$$\begin{aligned}
I_2(y_0) &= 4(\alpha-1/2)^2 \cdot \int_{1>z_1>z_2,z_0>0} P(y_0, y_1(z_1), y_2(z_2)) z_1^{2\alpha-2} z_2^{2\alpha-2} \, \mathrm{d}z_2 \, \mathrm{d}z_1\\
&= 4(\alpha-1/2)^2 \cdot \Bigg(\int_{1>z_1>z_0,z_2>0} z_0 z_1^{2\alpha-3} z_2^{2\alpha-2} \, \mathrm{d}z_2 \, \mathrm{d}z_1\\
&\quad - \int_{\substack{1>z_1>z_0,z_2>0\\ z_1<z_0+z_2}} G(z_1,z_0,z_2) z_0 z_1^{2\alpha-3} z_2^{2\alpha-2} \, \mathrm{d}z_2 \, \mathrm{d}z_1 \Bigg)\\
&=: 4(\alpha-1/2)^2 (I_{21}(y_0) - I_{22}(y_0)).
\end{aligned}$$

We start with the easy integral

$$\begin{aligned}
I_{21}(y_0) &= z_0 \int_{1>z_1>\max(z_2,z_0);z_0,z_2>0} z_1^{2\alpha-3} z_2^{2\alpha-2} \, \mathrm{d}z_2 \, \mathrm{d}z_1\\
&= z_0 \int_{z_0}^1 \int_0^{z_1} z_1^{2\alpha-3} z_2^{2\alpha-2} \, \mathrm{d}z_2 \, \mathrm{d}z_1\\
&= z_0 \int_{z_0}^1 \left[\frac{z_2^{2\alpha-1}}{2\alpha-1} \right]_0^{z_1} z_1^{2\alpha-3} \, \mathrm{d}z_1 = \frac{z_0}{2\alpha-1} \int_{z_0}^1 z_1^{4\alpha-4} \, \mathrm{d}z_1\\
&= \frac{z_0 - z_0^{4\alpha-2}}{(4\alpha-3)(2\alpha-1)}.
\end{aligned} \tag{4.16}$$

We note that the denominators above are non-zero as $\alpha > \frac{1}{2}$ and $\alpha \neq \frac{3}{4}$.

To deal with $I_{22}(y_0)$ we consider the function

$$J'_{a,b,c}(z_0) := z_0^a \int_{\substack{1>z_1>\max(z_0,z_2);z_0,z_2>0\\ z_1<z_0+z_2}} z_1^{b+2\alpha-2} z_2^{c+2\alpha-2} \, \mathrm{d}z_2 \, \mathrm{d}z_1$$

and write

$$
\begin{aligned}
I_{2,2}(y_0) = {} & \frac{1}{4}\left(J'_{0,-1,1}(z_0) + J'_{2,-1,-1}(z_0) + J'_{0,1,-1}(z_0)\right) \\
& + \frac{1}{2}\left(J'_{1,-1,0}(z_0) - J'_{0,0,0}(z_0) - J'_{1,0,-1}(z_0)\right).
\end{aligned}
\tag{4.17}
$$

We now compute $J'_{a,b,c}(z_0)$ as

$$
\begin{aligned}
J'_{a,b,c}(z_0) &= z_0^a \int_{z_0}^1 \int_{z_1-z_0}^{z_1} z_1^{b+2\alpha-2} z_2^{c+2\alpha-2} \, \mathrm{d}z_2 \, \mathrm{d}z_1 \\
&= z_0^a \int_{z_0}^1 \frac{1}{c+2\alpha-1} z_1^{b+2\alpha-2} (z_1^{c+2\alpha-1} - (z_1-z_0)^{c+2\alpha-1}) \, \mathrm{d}z_1 \\
&= z_0^a \int_{z_0}^1 \frac{1}{c+2\alpha-1} z_1^{b+c+4\alpha-3} \mathrm{d}z_1 \\
&\quad - z_0^a \int_{z_0}^1 \frac{1}{c+2\alpha-1} z_1^{b+2\alpha-2} (z_1-z_0)^{c+2\alpha-1} \, \mathrm{d}z_1 \\
&= z_0^a \frac{1}{(c+2\alpha-1)(b+c+4\alpha-2)} (1 - z_0^{b+c+4\alpha-2}) \\
&\quad - \frac{z_0^a}{c+2\alpha-1} z_0^{b+c+4\alpha-2} B^-(1-z_0; c+2\alpha, -b-c-4\alpha+2) \\
&= \frac{z_0^a - z_0^{4\alpha-2}}{(c+2\alpha-1)(b+c+4\alpha-2)} - \frac{z_0^{4\alpha-2} B^-(1-z_0; c+2\alpha, -b-c-4\alpha+2)}{c+2\alpha-1}.
\end{aligned}
$$

Here we have used that for $x \in \mathbb{R}, y > -1$ (note that $c+2\alpha-1 > -1$ as $c \geq -1$; at the fourth equality sign we use the substitution $t = \frac{u}{1-u}$ with $dt = (1-u)^{-2} du$):

$$
\begin{aligned}
\int_{z_0}^1 z_1^x (z_1-z_0)^y \, \mathrm{d}z_1 &= \int_0^{1-z_0} (s+z_0)^x s^y \, \mathrm{d}s \\
&= z_0^{x+y} \int_0^{1-z_0} ((s/z_0)+1)^x \, (s/z_0)^y \, \mathrm{d}s \\
&= z_0^{x+y+1} \int_0^{1/z_0-1} (t+1)^x t^y \, \mathrm{d}t \\
&= z_0^{x+y+1} \int_0^{1-z_0} u^y (1-u)^{-(x+y+2)} \, \mathrm{d}u \\
&= z_0^{x+y+1} B^-(1-z_0; y+1, -x-y-1).
\end{aligned}
$$

As $c \geq -1$ and $-a \in \{0,-1,-2\}$ and by our assumption $\alpha \notin \{\frac{3}{4}\}$, the denominators that occur during the above computations are non-zero.

Plugging the expression for $J'_{a,b,c}(z_0)$ back into (4.17) we get,

$$
I_{2,2}(y_0) = \frac{1 - z_0^{4\alpha-2}}{32\alpha(\alpha-1/2)} - \frac{z_0^{4\alpha-2} B^-(1-z_0; 1+2\alpha, -4\alpha+2)}{8\alpha}
$$

$$
\begin{aligned}
&+\frac{z_0^2 - z_0^{4\alpha-2}}{32(\alpha-1)^2} - \frac{z_0^{4\alpha-2} B^-(1-z_0; -1+2\alpha, -4\alpha+4)}{8(\alpha-1)} \\
&+\frac{1 - z_0^{4\alpha-2}}{32(\alpha-1)(\alpha-1/2)} - \frac{z_0^{4\alpha-2} B^-(1-z_0; -1+2\alpha, -4\alpha+2)}{8(\alpha-1)} \\
&+\frac{z_0 - z_0^{4\alpha-2}}{16(\alpha-1/2)(\alpha-3/4)} - \frac{z_0^{4\alpha-2} B^-(1-z_0; 2\alpha, -4\alpha+3)}{4(\alpha-1/2)} \\
&-\frac{1 - z_0^{4\alpha-2}}{16(\alpha-1/2)^2} + \frac{z_0^{4\alpha-2} B^-(1-z_0; 2\alpha, -4\alpha+2)}{4(\alpha-1/2)} \\
&-\frac{z_0 - z_0^{4\alpha-2}}{16(\alpha-1)(\alpha-3/4)} + \frac{z_0^{4\alpha-2} B^-(1-z_0; -1+2\alpha, -4\alpha+3)}{4(\alpha-1)}.
\end{aligned}
$$

Using some algebra and the identities (4.13) and (4.14) this can be reduced to

$$
\begin{aligned}
&I_{2,2}(y_0) \\
&= \frac{1}{64\alpha(\alpha-1/2)^2(\alpha-1)} - \frac{(1-z_0)^{2\alpha}}{64\alpha(\alpha-1/2)^2(\alpha-1)} - \frac{z_0}{8(\alpha-1/2)(\alpha-1)(4\alpha-3)} \\
&\quad + \frac{z_0^2}{32(\alpha-1)^2} + \frac{(-6+25\alpha-48\alpha^2+44\alpha^3-16\alpha^4) z_0^{4\alpha-2}}{512\alpha(\alpha-1/2)^2(\alpha-1)^2(\alpha-3/4)} \\
&\quad + \frac{z_0^{4\alpha-2} B^-(1-z_0; 2\alpha, 3-4\alpha)}{32(\alpha-1)(\alpha-1/2)^2}.
\end{aligned}
\tag{4.18}
$$

Combining the results for $I_1(y_0)$ and $I_2(y_0)$ Combining the results for $I_{11}(y_0), I_{12}(y_0), I_{21}(y_0)$ and $I_{22}(y_0)$ we get, after some algebra, an explicit expression for $P(y_0)$ as a linear combination of terms of the form z_0^u, $(1-z_0)^u$ and $z_0^u B^-(1-z_0; a, b)$ with $u \in \{0, 1, 2, 4\alpha-2, 2\alpha\}$:

$$
\begin{aligned}
&P(y_0) = 2(I_1 + I_2) = 8(\alpha-1/2)^2 (I_{1,1} - I_{1,2} + I_{2,1} - I_{2,2}) \\
&= 8(\alpha-1/2)^2 \Bigg(\frac{1}{2(2\alpha-1)^2} z_0^{4\alpha-2} \\
&\quad - \frac{\left(\frac{3}{64} - \frac{3}{16} 2^{-4\alpha} + \alpha(-\frac{41}{128} + \frac{13}{16} 2^{-4\alpha}) + \alpha^2(\frac{5}{8} - \frac{3}{4} 2^{-4\alpha}) - \frac{15}{32}\alpha^3 + \frac{1}{8} a^4\right) z_0^{4\alpha-2}}{4(\alpha-1/2)^2(\alpha-1)^2(\alpha-3/4)\alpha} \\
&\quad + \frac{z_0^{4\alpha-2} B^-(1/2; 2\alpha+1, 2\alpha-2)}{8(\alpha-1)\alpha(2\alpha-1)} + \frac{z_0 - z_0^{4\alpha-2}}{(4\alpha-3)(2\alpha-1)} \\
&\quad - \frac{1}{64\alpha(\alpha-1/2)^2(\alpha-1)} + \frac{(1-z_0)^{2\alpha}}{64\alpha(\alpha-1/2)^2(\alpha-1)} + \frac{z_0}{8(\alpha-1/2)(\alpha-1)(4\alpha-3)} \\
&\quad - \frac{z_0^2}{32(\alpha-1)^2} - \frac{(-6+25\alpha-48\alpha^2+44\alpha^3-16\alpha^4) z_0^{4\alpha-2}}{512\alpha(\alpha-1/2)^2(\alpha-1)^2(\alpha-3/4)} \\
&\quad - \frac{z_0^{4\alpha-2} B^-(1-z_0; 2\alpha, 3-4\alpha)}{32(\alpha-1)(\alpha-1/2)^2} \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{8(\alpha-1)\alpha} + \frac{(\alpha-1/2)z_0}{\alpha-1} - \frac{(\alpha-1/2)^2 z_0^2}{4(\alpha-1)^2} \\
&+ z_0^{-2+4\alpha}\left(\frac{2^{-4\alpha-1}(3\alpha-1)}{\alpha(\alpha-1)^2} + \frac{(\alpha-1/2)B^-(1/2;1+2\alpha,-2+2\alpha)}{2(\alpha-1)\alpha}\right) \\
&+ \frac{(1-z_0)^{2\alpha}}{8(\alpha-1)\alpha} - \frac{z_0^{4\alpha-2}B^-(1-z_0;2\alpha,3-4\alpha)}{4(\alpha-1)}.
\end{aligned}
$$

Observe that the above expression only contains terms of the form $\alpha - 1$ in the denominator. The only expression of the form $\alpha - 3/4$ is in the lower incomplete beta-function $B^-(1-z_0; 2\alpha, 3-4\alpha)$ which appears twice in the expression for $P(y_0)$.

The case of $\alpha = 3/4$

Note that the factor $\alpha - \frac{3}{4}$ does not occur in any denominator of the previously obtained expression. For the lower incomplete beta function, the last argument $3 - 4\alpha$ is zero for $\alpha = \frac{3}{4}$, however as $z_0 < 1$ the integration domain of the lower incomplete beta function does not touch the singularity at $t = 1$ (note $B^-(1 - z_0; 2\alpha, 3-4\alpha) = \int_0^{1-z_0} t^{2\alpha-1}(1-t)^{2-4\alpha}dt$). Therefore, the previous expression holds for this case as well.

Computing γ and $\gamma(k)$

Now that we have an expression for $P(y_0)$ we can compute $\gamma, \gamma(k)$ by integrating over y_0 and prove that they equal the expressions given in, respectively, Theorem 1.5.1 and Theorem 1.5.2.

We define

$$
I^{(k)} := \int_0^\infty P(y)\alpha e^{-\alpha y}\rho(y,k)\,\mathrm{d}y = \int_0^\infty P(y)\alpha e^{-\alpha y}\frac{\left(\xi e^{y/2}\right)^k}{k!}e^{-\xi e^{y/2}}\,\mathrm{d}y
$$

and

$$
J := \int_0^\infty P(y)\alpha e^{-\alpha y}\,\mathrm{d}y.
$$

Then, recalling (4.7) and (4.6), we have

$$
\gamma = J - I^{(1)} - I^{(2)} \quad \text{and} \quad \gamma(k) = \frac{I^{(k)}}{p_k}.
$$

We will thus compute J and $I^{(k)}$. It will be helpful to change coordinates to $z := e^{-y/2}$. This yields

$$
J = 2\alpha\int_0^1 P(y(z))z^{2\alpha-1}\,\mathrm{d}z,
$$

and

$$I^{(k)} = \frac{2\alpha\xi^k}{k!} \cdot \int_0^1 P(y(z)) \cdot z^{2\alpha-(k+1)} e^{-\xi z^{-1}} \, \mathrm{d}z.$$

We shall be assuming that $\alpha \neq 1$. We observe from Lemma 4.1.1 that for $\alpha \neq 1$, $P(y(z))$ is in fact a linear combination of terms of the form z^u, $(1-z)^u$ and $z^u B^-(1-z; v, w)$ with $u \in \{0, 1, 2, 4\alpha - 2, 2\alpha\}$.

To compute J we observe that, by integration by parts,

$$\begin{aligned}
&\int_0^1 z^{u+2\alpha-1} B^-(1-z; v, w) \, \mathrm{d}z \\
&= \left[\frac{z^{u+2\alpha}}{u+2\alpha} B^-(1-z; v, w) \right]_0^1 + \frac{1}{u+2\alpha} \int_0^1 z^{u+2\alpha+w-1}(1-z)^{v-1} \, \mathrm{d}z \\
&= \frac{1}{u+2\alpha} B(u+w+2\alpha, v),
\end{aligned}$$

where we have used that $\frac{\partial}{\partial z} B^-(1-z; v, w) = -z^{w-1}(1-z)^{v-1}$. This takes care of the two integrands involving the beta function in $P(y)$. The other integrals are easily computed and yield the following expression for J (note that it only depends on α but not on ν)

$$\begin{aligned}
J = {} & \frac{2 + 4\alpha + 13\alpha^2 - 34\alpha^3 - 12\alpha^4 + 24\alpha^5}{16(\alpha-1)^2 \alpha(\alpha+1)(2\alpha+1)} + \frac{2^{-1-4\alpha}}{(\alpha-1)^2} \\
& + \frac{(\alpha - 1/2)(B(2\alpha, 2\alpha+1) + B^-(1/2; 1+2\alpha, -2+2\alpha))}{2(\alpha-1)(3\alpha-1)}.
\end{aligned}$$

We proceed to work out $I^{(k)}$. For this we will compute the integrals involving terms in $P(y(z))$ of the form z^u, $(1-z)^u$ and $B(1-z, v, w)$ separately. We first point out that for any $0 \leq a < b \leq 1$

$$\begin{aligned}
\int_a^b z^{u+2\alpha-(k+1)} e^{-\xi z^{-1}} \, \mathrm{d}z &= \xi^{u+2\alpha-k} \int_{\xi/b}^{\xi/a} t^{k-1-2\alpha-u} e^{-t} \, \mathrm{d}t \\
&= \xi^{u+2\alpha-k} \left(\Gamma^+(k-2\alpha-u, \xi/b) - \Gamma^+(k-2\alpha-u, \xi/a) \right).
\end{aligned}$$

In particular

$$\int_0^1 z^{u+2\alpha-k-1} e^{-\xi z^{-1}} \, \mathrm{d}z = \xi^{u+2\alpha-k} \Gamma^+(k-2\alpha-u, \xi), \tag{4.19}$$

where Γ^+ denotes the (upper) incomplete gamma function, and we have used the substitution $t = \xi/z$ which gives $\mathrm{d}z = -\xi t^{-2} \, \mathrm{d}t$. (And of course it is understood that $\xi/0 = \infty$). This takes care of the integrals of all terms in $P(y(z))$ of the form z^u.

Next we will consider the integrals over the terms in $P(y(z))$ of the form $(1-z)^u$. For this we need the hypergeometric U-function (also called Tricomi's confluent hypergeometric function), which has the integral representation

$$U(a,b,z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1} \, \mathrm{d}t.$$

which holds for $a,b,z \in \mathbb{C}$, $b \notin \mathbb{Z}_{\leq 0}$, $Re(a), Re(z) > 0$, see [23, p.255]. Applying the change of variables $t = \frac{1-s}{s}$ (i.e. $\mathrm{d}t = -s^{-2}\,\mathrm{d}s$ and $s = \frac{1}{t+1}$) yields

$$U(a,b,z) = \frac{e^z}{\Gamma(a)} \int_0^1 s^{-b}(1-s)^{a-1} e^{-z/s} ds$$

Setting $a = 2\alpha + 1 > 0$, $b = -2\alpha + k + 1$, $z = \xi > 0$, then gives

$$\int_0^1 z_0^{2\alpha-k-1} e^{-\xi/z_0} (1-z_0)^{2\alpha} dz_0 = \Gamma(2\alpha+1) e^{-\xi} U(2\alpha+1, 1+k-2\alpha, \xi). \quad (4.20)$$

Finally we need to deal with the terms in $P(y(z))$ that involve the incomplete beta function. Let $a,c \in \mathbb{R}$, $\xi, b > 0$ positive real numbers. Using the integral definition of the incomplete beta function, the change of variables $s = 1-t$ gives:

$$\begin{aligned}\int_0^1 z^a e^{-\xi/z} B^-(1-z;b,c)\,\mathrm{d}z &= \int_0^1 z^a e^{-\xi/z} \int_0^{1-z} t^{b-1}(1-t)^{c-1}\,\mathrm{d}t\,\mathrm{d}z \\ &= \int_0^1 z^a e^{-\xi/z} \int_z^1 s^{c-1}(1-s)^{b-1}\,\mathrm{d}s\,\mathrm{d}z.\end{aligned}$$

Then changing the order of integration and using the substitution $u = \xi/z$ and recognizing the upper incomplete gamma function yields

$$\begin{aligned}&\int_0^1 z^a e^{-\xi/z} \int_z^1 s^{c-1}(1-s)^{b-1}\,\mathrm{d}s\,\mathrm{d}z \\ &\quad= \int_0^1 \int_0^s z^a e^{-\xi/z}\,\mathrm{d}z\, s^{c-1}(1-s)^{b-1}\,\mathrm{d}s \\ &\quad= \int_0^1 \int_{\xi/s}^\infty \xi^{a+1} u^{-a-2} e^{-u}\,\mathrm{d}u\, s^{c-1}(1-s)^{b-1}\,\mathrm{d}s \\ &\quad= \xi^{a+1} \int_0^1 \Gamma^+(-a-1, \xi/s) s^{c-1}(1-s)^{b-1}\,\mathrm{d}s. \qquad (4.21)\end{aligned}$$

To compute this last integral we make use of the fact that the incomplete Γ-function has a representation in terms of Meijer's G-function (see Lemma A.1 in Appendix A)

$$\Gamma^+(-a-1, \xi/s) = G^{2,0}_{1,2}\left(\frac{\xi}{s}\middle| \begin{matrix} 1 \\ -a-1, 0 \end{matrix}\right),$$

which holds for any $a \in \mathbb{R}$ and $s > 0$ (that for a fixed second argument, the upper incomplete gamma function is entire in the first argument, see [28, pp. 899, 1032ff.]). We can now evaluate the integral in (4.21) using several identities for Meijer's G-function. First, inserting the expression for the incomplete Gamma-function into (4.21) gives

$$\xi^{a+1} \int_0^1 s^{c-1}(1-s)^{b-1} G_{1,2}^{2,0}\left(\frac{\xi}{s}\middle|\begin{matrix}1\\-a-1,0\end{matrix}\right) \mathrm{d}s.$$

Next we apply the inversion identity for Meijer's G-function (see [23, p. 209, 5.3.1.(9))]) to get

$$\xi^{a+1} \int_0^1 s^{c-1}(1-s)^{b-1} G_{2,1}^{0,2}\left(\frac{s}{\xi}\middle|\begin{matrix}2+a,1\\0\end{matrix}\right) \mathrm{d}s.$$

This expression is actually the Euler transform of Meijer's G-function (see [23, p. 214, 5.5.2.(5)]) and (as the conditions $2+1 < 2(0+2)$ and $|\arg(\xi^{-1})| < \frac{\pi}{2}$ (as $\xi > 0$) and $1-c-b < 1-c$ (as $b > 0$) are satisfied) it equals

$$\xi^{a+1}\Gamma(b) G_{3,2}^{0,3}\left(\xi^{-1}\middle|\begin{matrix}1-c,2+a,1\\0,1-c-b\end{matrix}\right).$$

Using again the inversion identity for Meijer's G-function (see [23, p. 209, 5.3.1.(9))]) we now get

$$\xi^{a+1}\Gamma(b) G_{2,3}^{3,0}\left(\xi\middle|\begin{matrix}1,b+c\\c,-1-a,0\end{matrix}\right).$$

Finally, plugging in $a = 6\alpha - k - 3$, $b = 2\alpha$, $c = 3 - 4\alpha$ we obtain

$$\int_0^1 z^a e^{-\xi/z} B^-(1-z;b,c)\,\mathrm{d}z = \xi^{6a-k-2}\Gamma(2\alpha) G_{2,3}^{3,0}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\3-4\alpha,-6\alpha+k+2,0\end{matrix}\right). \tag{4.22}$$

Using (4.19), (4.20) and (4.22) we get

$$\begin{aligned}
I^{(k)} = \frac{\xi^{2\alpha}}{4k!(\alpha-1)} \Bigg(& -\Gamma^+(k-2\alpha,\xi) - 2\frac{\alpha(\alpha-1/2)^2\xi^2\Gamma^+(k-2\alpha-2,\xi)}{(\alpha-1)} \\
& +8\alpha(\alpha-1/2)\xi\Gamma^+(k-2\alpha-1,\xi) \\
& +4\xi^{4\alpha-2}\Gamma^+(k-6\alpha+2,\xi)\left(\frac{2^{-4\alpha}(3\alpha-1)}{(\alpha-1)} + (\alpha-1/2)B^-(1/2;1+2\alpha,-2+2\alpha)\right) \\
& +\xi^{k-2\alpha}\Gamma(2\alpha+1)e^{-\xi}U(2\alpha+1,1+k-2\alpha,\xi) \\
& -\xi^{4\alpha-2}\Gamma(2\alpha+1)G_{2,3}^{3,0}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\3-4\alpha,-6\alpha+k+2,0\end{matrix}\right)\Bigg).
\end{aligned}$$

With the expressions for J and $I^{(k)}$ and using $\Gamma^*(q,z) = \Gamma^+(q+1,z)+\Gamma^+(q,z)$ we now obtain, after some algebra,

$$
\begin{aligned}
\gamma &= J - I^{(0)} - I^{(1)} \\
&= \frac{2+4\alpha+13\alpha^2-34\alpha^3-12\alpha^4+24\alpha^5}{16(\alpha-1)^2\alpha(\alpha+1)(2\alpha+1)} + \frac{2^{-1-4\alpha}}{(\alpha-1)^2} \\
&\quad + \frac{(\alpha-1/2)(B(2\alpha,2\alpha+1)+B^-(1/2;1+2\alpha,-2+2\alpha))}{2(\alpha-1)(3\alpha-1)} \\
&\quad - \frac{\xi^{2\alpha}}{4(\alpha-1)}\Bigg(-\Gamma^+(-2\alpha,\xi) - 2\frac{\alpha(\alpha-1/2)^2\xi^2\Gamma^+(-2\alpha-2,\xi)}{(\alpha-1)} \\
&\qquad +8\alpha(\alpha-1/2)\xi\Gamma^+(-2\alpha-1,\xi) \\
&\qquad +4\xi^{4\alpha-2}\Gamma^+(-6\alpha+2,\xi)\left(\frac{2^{-4\alpha}(3\alpha-1)}{(\alpha-1)} + (\alpha-1/2)B^-(1/2;1+2\alpha,-2+2\alpha)\right) \\
&\qquad +\xi^{-2\alpha}\Gamma(2\alpha+1)e^{-\xi}U(2\alpha+1,1-2\alpha,\xi) \\
&\qquad -\xi^{4\alpha-2}\Gamma(2\alpha+1)G^{3,0}_{2,3}\left(\xi\,\middle|\,\begin{matrix}1,3-2\alpha\\ 3-4\alpha,-6\alpha+2,0\end{matrix}\right)\Bigg) \\
&\quad - \frac{\xi^{2\alpha}}{4(\alpha-1)}\Bigg(-\Gamma^+(1-2\alpha,\xi) - 2\frac{\alpha(\alpha-1/2)^2\xi^2\Gamma^+(-2\alpha-1,\xi)}{(\alpha-1)} \\
&\qquad +8\alpha(\alpha-1/2)\xi\Gamma^+(1-2\alpha-1,\xi) \\
&\qquad +4\xi^{4\alpha-2}\Gamma^+(1-6\alpha+2,\xi)\left(\frac{2^{-4\alpha}(3\alpha-1)}{(\alpha-1)} + (\alpha-1/2)B^-(1/2;1+2\alpha,-2+2\alpha)\right) \\
&\qquad +\xi^{1-2\alpha}\Gamma(2\alpha+1)e^{-\xi}U(2\alpha+1,2-2\alpha,\xi) \\
&\qquad -\xi^{4\alpha-2}\Gamma(2\alpha+1)G^{3,0}_{2,3}\left(\xi\,\middle|\,\begin{matrix}1,3-2\alpha\\ 3-4\alpha,-6\alpha+3,0\end{matrix}\right)\Bigg) \\
&= \frac{2+4\alpha+13\alpha^2-34\alpha^3-12\alpha^4+24\alpha^5}{16(\alpha-1)^2\alpha(\alpha+1)(2\alpha+1)} + \frac{2^{-1-4\alpha}}{(\alpha-1)^2} \\
&\quad + \frac{(\alpha-1/2)(B(2\alpha,2\alpha+1)+B^-(1/2;1+2\alpha,-2+2\alpha))}{2(\alpha-1)(3\alpha-1)} \\
&\quad + \frac{\xi^{2\alpha}\Gamma^*(-2\alpha,\xi)}{4(\alpha-1)} + \frac{\xi^{2\alpha+2}\alpha(\alpha-1/2)^2\Gamma^*(-2\alpha-2,\xi)}{2(\alpha-1)^2} \\
&\quad - \frac{\xi^{2\alpha+1}\alpha(2\alpha-1)\Gamma^*(-2\alpha-1,\xi)}{(\alpha-1)} - \frac{\xi^{6\alpha-2}2^{-4\alpha}(3\alpha-1)\Gamma^*(-6\alpha+2,\xi)}{(\alpha-1)^2} \\
&\quad - \frac{\xi^{6\alpha-2}(\alpha-1/2)B^-(1/2;1+2\alpha,-2+2\alpha)\Gamma^*(-6\alpha+2,\xi)}{(\alpha-1)} \\
&\quad - \frac{e^{-\xi}\Gamma(2\alpha+1)\left(U(2\alpha+1,1-2\alpha,\xi)+U(2\alpha+1,2-2\alpha,\xi)\right)}{4(\alpha-1)}
\end{aligned}
$$

$$+\frac{\xi^{6\alpha-2}\Gamma(2\alpha+1)\left(G^{3,0}_{2,3}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\3-4\alpha,-6\alpha+2,0\end{matrix}\right)+G^{3,0}_{2,3}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\3-4\alpha,-6\alpha+3,0\end{matrix}\right)\right)}{4(\alpha-1)},$$

which is the expression in Theorem 1.5.1.

Similarly, we get

$$\begin{aligned}
\gamma(k) &= \frac{I^{(k)}}{p_k}\\
&= \frac{1}{8\alpha(\alpha-1)\Gamma^+(k-2\alpha,\xi)}\Bigg(-\Gamma^+(k-2\alpha,\xi)-2\frac{\alpha(\alpha-1/2)^2\xi^2\Gamma^+(k-2\alpha-2,\xi)}{(\alpha-1)}\\
&\quad+8\alpha(\alpha-1/2)\xi\Gamma^+(k-2\alpha-1,\xi)\\
&\quad+4\xi^{4\alpha-2}\Gamma^+(k-6\alpha+2,\xi)\left(\frac{2^{-4\alpha}(3\alpha-1)}{(\alpha-1)}+(\alpha-1/2)B^-(1/2;1+2\alpha,-2+2\alpha)\right)\\
&\quad+\xi^{k-2\alpha}\Gamma(2\alpha+1)e^{-\xi}U(2\alpha+1,1+k-2\alpha,\xi)\\
&\quad-\xi^{4\alpha-2}\Gamma(2\alpha+1)G^{3,0}_{2,3}\left(\xi\middle|\begin{matrix}1,3-2\alpha\\3-4\alpha,-6\alpha+k+2,0\end{matrix}\right)\Bigg),
\end{aligned}$$

which equals the expression in Theorem 1.5.2.

Explicit expressions for $\gamma, \gamma(k)$ when $\alpha = 1$.

Although we have already established that $\gamma, \gamma(k)$ can be obtained at $\alpha = 1$ by taking the $\alpha \to 1$ limit of the expression obtained for $\alpha = 1$, it is still helpful to derive an alternative, more explicit expression. This is what we will do in the current section. We will prove

Proposition 4.1.5. *If $\alpha = 1$ then*

$$\begin{aligned}
\gamma &= \frac{575-12\pi^2}{576}+\frac{\eta^4(7+\pi^2)\left(\Gamma^+(-3,\eta)+\Gamma^+(-4,\eta)\right)}{4}\\
&\quad-\frac{1}{2}\int_0^1(1-4z+3z^3)\log(1-z)(z+\eta)e^{-\eta/z}\,\mathrm{d}z\\
&\quad-\int_0^1\mathrm{Li}_2(z)(z^3+\eta z^2)e^{-\eta/z}\,\mathrm{d}z,
\end{aligned}$$

and

$$\begin{aligned}
\gamma(k) &= \frac{9\eta^3}{2k!}\Gamma^+(k-3,\eta)-\frac{\xi^4}{k!}\frac{7+\pi^2}{4}\Gamma^+(k-4,\eta)\\
&\quad+\frac{\eta^k}{2k!}\int_0^1(1-4z+3z^2)\ln(1-z)z^{1-k}e^{-\eta/z}\,\mathrm{d}z\\
&\quad+\frac{\eta^k}{k!}\int_0^1 z^{3-k}\,\mathrm{Li}_2(z)e^{-\eta/z}\,\mathrm{d}z,
\end{aligned}$$

with $\eta = \frac{4\nu}{\pi}$ *and* $\mathrm{Li}_2(z) = \sum_{t=1}^{\infty} \frac{z^t}{t^2} = \int_0^z \frac{\ln(1-t)}{t} dt$, *the dilogarithm function (which is a special case of the polylogarithm).*[1]

Naturally, the proof proceeds by proving the analogue of Lemma 4.1.1:

Lemma 4.1.6. *If* $\alpha = 1$, *then*

$$P(y_0) = \frac{9}{4} e^{-\frac{1}{2} y_0} + \frac{1 - 4e^{-\frac{1}{2} y_0} + 3e^{-y_0}}{4} \ln(1 - e^{-\frac{1}{2} y_0}) \\ - \frac{7 + \pi^2}{8} e^{-y_0} + \frac{1}{2} e^{-y_0} \mathrm{Li}_2(e^{-y_0}),$$

where $\mathrm{Li}_2(z)$ *is the dilogarithm function.*

Proof. We want to compute the limit $\lim_{\alpha \to 1} P_\alpha(y_0(z_0))$. For $\alpha \neq 1$, we label the terms as follows:

$$P_\alpha(y_0(z_0)) = \frac{1}{\alpha - 1} \Bigg(s_1(\alpha, z_0) + s_2(\alpha, z_0) \\ + \frac{1}{\alpha - 1} (s_3(\alpha, z_0) + s_4(\alpha, z_0)) + s_5(\alpha, z_0) + s_6(\alpha, z_0) + s_7(\alpha, z_0) \Bigg),$$

where

$$\begin{aligned} s_1(\alpha, z_0) &= -\frac{1}{8\alpha}, \\ s_2(\alpha, z_0) &= (\alpha - 1/2) z_0, \\ s_3(\alpha, z_0) &= -\frac{(\alpha - 1/2)^2 z_0^2}{4}, \\ s_4(\alpha, z_0) &= z_0^{-2+4\alpha} \frac{2^{-4\alpha-1}(3\alpha - 1)}{\alpha}, \\ s_5(\alpha, z_0) &= z_0^{-2+4\alpha} \frac{(\alpha - 1/2) B^-(1/2; 1 + 2\alpha, -2 + 2\alpha)}{2\alpha}, \\ s_6(\alpha, z_0) &= \frac{(1 - z_0)^{2\alpha}}{8\alpha}, \\ s_7(\alpha, z_0) &= -\frac{z_0^{4\alpha-2} B^-(1 - z_0; 2\alpha, 3 - 4\alpha)}{4}. \end{aligned}$$

Now, we consider the functions $s_i(\alpha) = s_i(\alpha, z_0)$ as functions of α only and compute their Taylor expansion at $\alpha = 1$, for $i \in \{1, 2, 5, 6, 7\}$ up to linear order and for

[1] Note that the integrals in the expression for γ for $\alpha = 1$ exist: for the first one note that $1 - 4z + 3z^2 = (1 - z)(1 - 3z)$, so the integrand can be bounded by $C(1 - z)\log(1 - z)$ on $[0, 1)$ for some constant C, which can be continued continuously to the compact interval $[0, 1]$ by noting that the limit for $z \to 1$ is zero, so the integrand is bounded on a bounded domain and hence, this integral is finite; for the second integral note that $\mathrm{Li}_2(z)$ is bounded by $\mathrm{Li}_2(1)$ on $[0, 1]$, which is a series with well-known finite limit, so again the integrand is bounded on a bounded domain and hence the second integral is also finite.

$i \in \{3,4\}$ up to quadratic order, i.e. we write $s_i(\alpha) = s_i(1)+s_i'(1)(\alpha-1)+o(\alpha-1)$ for $i \in \{1,2,5,6,7\}$ and $s_i(\alpha) = s_i(1) + s_i'(1)(\alpha-1) + \frac{s_i''(1)}{2}(\alpha-1)^2 + o((\alpha-1)^2)$ for $i \in \{3,4\}$. Using these expansions, we can rewrite

$$P(y_0(z_0)) = \frac{1}{\alpha-1}\left(\sum_{i\in\{1,2,5,6,7\}} s_i(1) + \sum_{i\in\{1,2,5,6,7\}} s_i'(1)(\alpha-1) + o(\alpha-1)\right.$$
$$\left. + \frac{s_3(1)+s_4(1)}{\alpha-1} + s_3'(1) + s_4'(1) + \frac{1}{2}(s_3''(1)+s_4''(1))(\alpha-1) + o((\alpha-1))\right).$$

In order to continue, we compute:

$$s_1(\alpha) = -\frac{1}{8} + \frac{1}{8}(\alpha-1) + o(\alpha-1),$$
$$s_2(\alpha) = \frac{1}{2}z_0 + z_0(\alpha-1) + o(\alpha-1),$$
$$s_3(\alpha) = -\frac{1}{16}z_0^2 - \frac{1}{4}z_0^2(\alpha-1) - \frac{1}{2}z_0^2(\alpha-1)^2 + o((\alpha-1)^2),$$
$$s_4(\alpha) = \frac{1}{16}z_0^2 + \frac{z_0^2}{4}\left(\frac{1}{8} + \ln\frac{z_0}{2}\right)(\alpha-1)$$
$$+ \frac{z_0^2}{8}\left(8\left(\ln\frac{z_0}{2}\right)^2 + 2\ln\frac{z_0}{2} - \frac{1}{2}\right)(\alpha-1)^2 + o((\alpha-1)^2),$$
$$s_5(\alpha) = \frac{z_0^2}{4}B^-(1/2;3,0) + o(\alpha-1)$$
$$+ z_0^2\left(\left(\ln(z_0) + \frac{1}{4}\right)B^-(1/2;3,0) + 1/2\int_0^{\frac{1}{2}} \ln(t(1-t))t^2(1-t)^{-1}\,\mathrm{d}t\right)(\alpha-1),$$
$$s_6(\alpha) = \frac{(1-z_0)^2}{8} + \frac{(1-z_0)^2}{4}(\ln(1-z_0) - 1/2)(\alpha - 1 + o\,(a-1)),$$
$$s_7(\alpha) = -\frac{z_0^2}{4}B^-(1-z_0;2,-1) + o(\alpha-1)$$
$$- z_0^2\left(\ln(z_0)B^-(1-z_0;2,-1) + \int_0^{1-z_0} t(1-t)^{-2}\ln\left(\frac{\sqrt{t}}{1-t}\right)t(1-t)^{-2}\,\mathrm{d}t\right)(\alpha-1).$$

Based on this, we see that

$$s_3(1) + s_4(1) = -\frac{1}{16}z_0^2 + \frac{1}{16}z_0^2 = 0,$$

and

$$\sum_{i\in\{1,2,5,6,7\}} s_i(1) + s_3'(1) + s_4'(1)$$
$$= -\frac{1}{8} + \frac{1}{2}z_0 - \frac{1}{4}z_0^2 + \frac{z_0^2}{32} + \frac{z_0^2}{4}\ln(\frac{z_0}{2}) + \frac{z_0^2}{4}B^-(1/2;3,0)$$

$$
\begin{aligned}
&\quad + \frac{(1-z_0)^2}{8} - \frac{z_0^2}{4} B^-(1-z_0;2,-1) \\
&= -\frac{1}{8} + \frac{1}{2}z_0 - \frac{1}{4}z_0^2 + \frac{z_0^2}{32} + \left(\frac{z_0^2}{4}\ln(z_0) - \frac{z_0^2}{4}\ln 2\right) + \left(-\frac{5z_0^2}{32} + \frac{z_0^2}{4}\ln 2\right) \\
&\quad + \left(\frac{1}{8} - \frac{z_0}{4} + \frac{z_0^2}{8}\right) + \left(\frac{z_0^2}{4} - \frac{z_0}{4} - \frac{z_0^2}{4}\ln z_0\right) \\
&= 0,
\end{aligned}
$$

using that

$$
\begin{aligned}
B^-(\frac{1}{2};3,0) &= \int_0^{\frac{1}{2}} t^2(1-t)^-1dt = \int_{\frac{1}{2}}^1 (1-s)^2 s^{-1} ds \\
&= \int_{\frac{1}{2}}^1 s^{-1} - 2 + sds = -2 + \frac{1}{2} - \ln\frac{1}{2} + 1 - \frac{1}{8} = -\frac{5}{8} + \ln 2,
\end{aligned}
$$

and

$$
\begin{aligned}
B^-(1-z_0;2,-1) &= \int_0^{1-z_0} t(1-t)^{-2}dt = \int_{z_0}^1 (1-s)s^{-2}ds \\
&= \int_{z_0}^1 s^{-2} - s^{-1}ds = -1 + z_0^{-1} + \ln z_0.
\end{aligned}
$$

Finally, it follows that as $\alpha \to 1$,

$$
P(y_0(z_0)) = \sum_{i\in\{1,2,5,6,7\}} s_i'(1) + \frac{1}{2}(s_3''(1) + s_4''(1)) + o(1).
$$

Therefore, the desired value of $\lim_{\alpha\to 1} P(y_0(z_0))$ is given by

$$
\begin{aligned}
&\sum_{i\in\{1,2,5,6,7\}} s_i'(1) + \frac{1}{2}(s_3''(1) + s_4''(1)) \\
&= \frac{1}{8} + z_0 - \frac{z_0^2}{4} + \frac{z_0^2}{8}(4(\ln\frac{z_0}{2})^2 + \ln\frac{z_0}{2} - \frac{1}{4}) + \frac{(1-z_0)^2}{4}(\ln(1-z_0) - 1/2) \\
&\quad + z_0^2\left(\left(\ln(z_0) + \frac{1}{4}\right) B^-(1/2;3,0) + \frac{1}{2}\int_0^{\frac{1}{2}} \ln(t(1-t))\, t^2(1-t)^{-1}\,\mathrm{d}t\right) \\
&\quad - z_0^2\left(\ln(z_0)B^-(1-z_0;2,-1) + \int_0^{1-z_0} \ln\left(\frac{\sqrt{t}}{1-t}\right) t(1-t)^{-2}\,\mathrm{d}t\right) \\
&= \frac{1}{8} + z_0 - \frac{z_0^2}{4} + \frac{z_0^2}{2}(\ln\frac{z_0}{2})^2 + \frac{z_0^2}{8}\ln\frac{z_0}{2} - \frac{z_0^2}{32} \\
&\quad - \frac{5}{8}z_0^2\ln(z_0) + z_0^2\ln(z_0)\ln 2 - \frac{5z_0^2}{32} + \frac{z_0^2\ln 2}{4}
\end{aligned}
$$

$$
\begin{aligned}
&+ z_0^2/2 \int_0^{\frac{1}{2}} \ln(t(1-t))\, t^2(1-t)^{-1}\, \mathrm{d}t \\
&+ \frac{(1-z_0)^2}{4} \ln(1-z_0) - \frac{1}{8} + \frac{z_0}{4} - \frac{z_0^2}{8} \\
&+ z_0^2 \ln(z_0) - z_0 \ln z_0 - z_0^2 (\ln z_0)^2 - z_0^2 \int_0^{1-z_0} \ln\left(\frac{\sqrt{t}}{1-t}\right) t(1-t)^{-2}\, \mathrm{d}t \\
= &\frac{5}{4} z_0 - \frac{9}{16} z_0^2 + \frac{z_0^2}{2} (\ln \frac{z_0}{2})^2 + \frac{z_0^2}{8} \ln \frac{z_0}{2} + \frac{(1-z_0)^2}{4} \ln(1-z_0) \\
&+ \frac{3}{8} z_0^2 \ln(z_0) + z_0^2 \ln(z_0) \ln 2 + \frac{z_0^2 \ln 2}{4} + z_0^2/2 \int_0^{\frac{1}{2}} \ln(t(1-t))\, t^2 (1-t)^{-1}\, \mathrm{d}t \\
&- z_0 \ln z_0 - z_0^2 (\ln z_0)^2 - z_0^2 \int_0^{1-z_0} \ln\left(\frac{\sqrt{t}}{1-t}\right) t(1-t)^{-2}\, \mathrm{d}t \\
= &\frac{5}{4} z_0 - \frac{9}{16} z_0^2 + \frac{z_0^2}{2} (\ln \frac{z_0}{2})^2 + \frac{z_0^2}{8} \ln \frac{z_0}{2} + \frac{(1-z_0)^2}{4} \ln(1-z_0) \\
&+ \frac{3}{8} z_0^2 \ln(z_0) + z_0^2 \ln(z_0) \ln 2 + \frac{z_0^2 \ln 2}{4} + \frac{z_0^2}{2} (\frac{11}{8} - \frac{1}{4} \ln 2 - \frac{3}{2} \ln(2)^2 - \mathrm{Li}_2(\frac{1}{2})) \\
&- z_0 \ln z_0 - z_0^2 (\ln z_0)^2 + z_0 \Bigg(1 + \frac{1}{2}(2-z_0) \ln(z_0) + \frac{1}{2} z_0 \ln(z_0)^2 \\
&- \frac{1}{2}(1-z_0) \ln(1-z_0) + \frac{1}{2} z_0 \mathrm{Li}_2(z_0)\Bigg) - z_0^2 - \frac{1}{2} z_0^2 \mathrm{Li}_2(1) \\
= &\frac{9}{4} z_0 - \frac{25}{16} z_0^2 + \frac{z_0^2}{2} (\ln \frac{z_0}{2})^2 + \frac{z_0^2}{8} \ln \frac{z_0}{2} + \frac{(1-z_0)^2}{4} \ln(1-z_0) \\
&- \frac{1}{8} z_0^2 \ln(z_0) + z_0^2 \ln(z_0) \ln 2 + \frac{z_0^2 \ln 2}{4} + z_0^2/2(11/8 - 1/4 \ln 2 - 3/2 \ln(2)^2 \\
&- \mathrm{Li}_2(1/2) - \mathrm{Li}_2(1) + \mathrm{Li}_2(z_0)) - \frac{1}{2} z_0^2 (\ln z_0)^2 - \frac{1}{2} z_0 (1-z_0) \ln(1-z_0),
\end{aligned}
$$

where we have used that

$$
\begin{aligned}
&z_0^2/2 \int_0^{\frac{1}{2}} \ln(t) t^2 (1-t)^{-1} + \ln(1-t) t^2 (1-t)^{-1}\, \mathrm{d}t \\
&= 11/8 - 1/4 \ln 2 - 3/2 \ln(2)^2 - \mathrm{Li}_2(1/2),
\end{aligned}
$$

and

$$
\begin{aligned}
&z_0^2 \int_0^{1-z_0} 1/2 \ln(t) t (1-t)^{-2} - t \ln(1-t)(1-t)^{-2}\, \mathrm{d}t \\
= &-\frac{1}{z_0} \left(1 + \frac{1}{2}(2-z_0) \ln(z_0) + \frac{1}{2} z_0 \ln(z_0)^2 - \frac{1}{2}(1-z_0) \ln(1-z_0) + \frac{1}{2} z_0 \mathrm{Li}_2(z_0)\right) \\
&+ 1 + \frac{1}{2} \mathrm{Li}_2(1).
\end{aligned}
$$

By expanding the squares and collecting terms, the last expression can be simplified to

$$\begin{aligned}
&\frac{9}{4}z_0 + \frac{1-4z_0+3z_0^2}{4}\ln(1-z_0) \\
&\qquad + z_0^2\left(-7/8 - \frac{\ln(2)^2 + 2\,\mathrm{Li}_2(1/2) + 2\,\mathrm{Li}_2(1)}{4}\right) + \frac{1}{2}z_0^2\,\mathrm{Li}_2(z) \\
&= \frac{9}{4}z_0 + \frac{1-4z_0+3z_0^2}{4}\ln(1-z_0) - \frac{7+\pi^2}{8}z_0^2 + \frac{1}{2}z_0^2\,\mathrm{Li}_2(z),
\end{aligned}$$

which finishes the computation. □

Proof of Proposition 4.1.5: It suffices to find the value of J and $I^{(k)}$ at $\alpha = 1$. We can do this by computing the integrals with the expression for $P(y)$ that we found for $\alpha = 1$, i.e.

$$\begin{aligned}
J &= 2\alpha\int_0^1 \left(\frac{9}{4}z + \frac{1-4z+3z^2}{4}\ln(1-z) - \frac{7+\pi^2}{8}z^2 + \frac{1}{2}z^2\,\mathrm{Li}_2(z)\right) z^{2\alpha-1}\,\mathrm{d}z \\
&= \frac{575-12\pi^2}{576},
\end{aligned}$$

and

$$\begin{aligned}
I^{(k)} &= \frac{2\alpha\xi^k}{k!}\int_0^1 \left(\frac{9}{4}z + \frac{1-4z+3z^2}{4}\ln(1-z) - \frac{7+\pi^2}{8}z^2 + \frac{1}{2}z^2\,\mathrm{Li}_2(z)\right) \\
&\qquad\qquad \times z^{2\alpha-k-1}e^{-\xi/z}\,\mathrm{d}z \\
&= \frac{2\eta^k}{k!}\int_0^1 \left(\frac{9}{4}z + \frac{1-4z+3z^2}{4}\ln(1-z) - \frac{7+\pi^2}{8}z^2 + \frac{1}{2}z^2\,\mathrm{Li}_2(z)\right) \\
&\qquad\qquad \times z^{1-k}e^{-\eta/z}\,\mathrm{d}z \\
&= \frac{9\eta^k}{2k!}\eta^{3-k}\Gamma^+(k-3,\eta) - \frac{\eta^k}{k!}\frac{7+\pi^2}{4}\eta^{4-k}\Gamma^+(k-4,\eta) \\
&\quad + \frac{\eta^k}{2k!}\int_0^1 (1-4z+3z^2)\ln(1-z)z^{1-k}e^{-\eta/z}dz + \frac{\eta^k}{k!}\int_0^1 z^{3-k}\,\mathrm{Li}_2(z)e^{-\eta/z}\,\mathrm{d}z \\
&= \frac{9\eta^3}{2k!}\Gamma^+(k-3,\eta) - \frac{\eta^4}{k!}\frac{7+\pi^2}{4}\Gamma^+(k-4,\eta) \\
&\quad + \frac{\eta^k}{2k!}\int_0^1 (1-4z+3z^2)\ln(1-z)z^{1-k}e^{-\eta/z}dz + \frac{\eta^k}{k!}\int_0^1 z^{3-k}\,\mathrm{Li}_2(z)e^{-\eta/z}\,\mathrm{d}z,
\end{aligned}$$

where $\eta = \frac{4\nu}{\pi}$ and and $\mathrm{Li}_2(z) = \sum_{t=1}^{\infty} z^t/t^2$, the dilogarithm function. Plugging this into (4.7) and (4.6) yields the expressions in the statement of the proposition. ■

4.1.3 The proof of Proposition 1.5.4

Instead of extracting the scaling of $\gamma(k)$ from its explicit expression, it turns out to be more convenient to derive it directly. Recall that

$$\gamma(k) = \frac{\int_0^\infty \rho(y,k)P(y)\alpha e^{-\alpha y}\,\mathrm{d}y}{\int_0^\infty \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}.$$

The asymptotic behaviour of the denominator follows from (4.4). Hence, the main term to consider is the numerator

$$\int_0^\infty P(y)\,\rho(y,k)\alpha e^{-\alpha y}\,dy,$$

and in particular the function $P(y)$. We therefore start with establishing the asymptotic behavior of the latter. First we combine (4.4) and (4.5) to obtain the following scaling result

$$\frac{\int_0^\infty e^{-\beta y}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}{\int_0^\infty \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y} \sim \xi^{2\beta}k^{-2\beta}. \tag{4.23}$$

Proposition 4.1.7 (Asymptotic behavior of $P(y)$)**.** *Let $\alpha > \frac{1}{2}$, $\nu > 0$ and $c_{\alpha,\nu}$ as defined in Proposition 1.5.4. Then, as $y \to \infty$,*

1. *for $\frac{1}{2} < \alpha < \frac{3}{4}$,*
$$P(y) \sim e^{-\frac{y}{2}(4\alpha-2)}c_{\alpha,\nu}\xi^{4\alpha-2},$$

2. *for $\alpha = \frac{3}{4}$,*
$$P(y) \sim \frac{y}{2}e^{-\frac{y}{2}},$$

3. *and for $\alpha > \frac{3}{4}$,*
$$P(y) \sim e^{-\frac{y}{2}}\frac{\alpha-\frac{1}{2}}{\alpha-\frac{3}{4}}.$$

Proof. We shall deal with each of the three cases for α separately.

Proof for $1/2 < \alpha < 3/4$ By Lemma 4.1.1 we get that

$$\begin{aligned} e^{(4\alpha-2)\frac{y}{2}}P(y) = {} & \frac{2^{-4\alpha-1}(3\alpha-1)}{\alpha(\alpha-1)^2} + \frac{(\alpha-\frac{1}{2})B^-(\frac{1}{2};1+2\alpha,-2+2\alpha)}{2(\alpha-1)\alpha} \\ & - \frac{B^-(1-e^{-\frac{y}{2}};2\alpha,3-4\alpha)}{4(\alpha-1)} \\ & + \frac{e^{(4\alpha-2)\frac{y}{2}}}{8(\alpha-1)\alpha}\Big((1-e^{-\frac{y}{2}})^{2\alpha}-1\Big) \end{aligned}$$

$$+\frac{\alpha-\frac{1}{2}}{\alpha-1}e^{(4\alpha-3)\frac{y}{2}}-\frac{(\alpha-\frac{1}{2})^2}{4(\alpha-1)^2}e^{4(\alpha-1)\frac{y}{2}}.$$

Because for any $b<1$, $B^-(1-z;a,b)$ converges to $B(a,b)<\infty$ as $z\to 0$, we get that as $y\to\infty$, the first three terms are asymptotically equivalent to

$$\begin{aligned}&\frac{3\alpha-1}{2^{4\alpha+1}\alpha(\alpha-1)^2}+\frac{(\alpha-\frac{1}{2})B^-(\frac{1}{2};1+2\alpha,-2+2\alpha)}{2(\alpha-1)\alpha}-\frac{B(2\alpha,3-4\alpha)}{4(\alpha-1)}\\&=c_{\alpha,\nu}\xi^{-(4\alpha-2)},\end{aligned}$$

where we recall that $c_{\alpha,\nu}$ was defined in Proposition 1.5.4 as

$$c_{\alpha,\nu}=\left(\frac{3\alpha-1}{2^{4\alpha+1}\alpha(\alpha-1)^2}+\frac{(\alpha-\frac{1}{2})B^-(\frac{1}{2};1+2\alpha,-2+2\alpha)}{2(\alpha-1)\alpha}-\frac{B(2\alpha,3-4\alpha)}{4(\alpha-1)}\right)\xi^{4\alpha-2}.$$

The proof now follows since for $1/2<\alpha<3/4$, the remaining three terms go to zero as $y\to\infty$.

Proof for $\boldsymbol{\alpha = 3/4}$ Similarly to the previous case, we use Lemma 4.1.1 to obtain (evaluating the expressions for $\alpha=3/4$)

$$\begin{aligned}\frac{2}{y}e^{\frac{y}{2}}P(y)=&\frac{2}{y}B^-(1-e^{-\frac{y}{2}};3/2,0)-\frac{4}{y}\frac{e^{\frac{y}{2}}\left((1-e^{-\frac{y}{2}})^{3/2}-1\right)}{3}-\frac{1}{y}-\frac{e^{-\frac{y}{2}}}{4y}\\&+\frac{2}{y}\left(\frac{5}{3}-\frac{2B^-(\frac{1}{2};5/2,-1/2)}{3}\right).\end{aligned}$$

First we note that as $y\to\infty$,

$$e^{\frac{y}{2}}\left(\left(1-e^{-\frac{y}{2}}\right)^{3/2}-1\right)\sim-\frac{3}{2},\tag{4.24}$$

which implies that

$$\lim_{y\to\infty}\frac{4}{y}\frac{e^{\frac{y}{2}}\left((1-e^{-\frac{y}{2}})^{3/2}-1\right)}{3}=0.$$

We can now conclude that all terms in $\frac{2}{y}e^{\frac{y}{2}}P(y)$ except the first one are $o(1)$ as $y\to\infty$. By writing $z=e^{-\frac{y}{2}}$ we can rewrite the first term as

$$\frac{2}{y}B^-(1-e^{-\frac{y}{2}};3/2,0)=-\frac{B^-(1-z;3/2,0)}{\log(z)}.$$

Since $B^-(1-z;3/2,0)\sim-\log(z)$ as $z\to 0$, see Lemma B.1, it now follows that for $\alpha=3/4$,

$$\lim_{y\to\infty}\frac{2}{y}B^-(1-e^{-\frac{y}{2}};3/2,0)=\lim_{z\to 0}-\frac{1}{\log(z)}B^-(1-z;3/2,0)=1.$$

We therefore conclude that

$$P(y)\sim\frac{y}{2}e^{-\frac{y}{2}},$$

as $y\to\infty$.

Proof for $\alpha > 3/4$ We first deal with the case $\alpha = 1$. Here it follows from Proposition 4.1.5 that

$$\begin{aligned} e^{y/2}P(y) &= \frac{9}{4} + \frac{e^{y/2}\log(1-e^{-y/2})}{4} \\ &\quad - \log(1-e^{-y/2}) + e^{-y/2}\left(\frac{3}{4}\log(1-e^{-y/2}) - \frac{7+\pi^2}{8} + \frac{1}{2}\mathrm{Li}_2(e^{-y})\right) \\ &= 2 + \left(\frac{e^{y/2}\log(1-e^{-y/2})}{4} + 1\right) \\ &\quad - \log(1-e^{-y/2}) + e^{-y/2}\left(\frac{3}{4}\log(1-e^{-y/2}) - \frac{7+\pi^2}{8} + \frac{1}{2}\mathrm{Li}_2(e^{-y})\right) \end{aligned}$$

The last two terms are $o(1)$ as $y \to \infty$, while $2 = (\alpha - 1/2)/(\alpha - 3/4)$ for $\alpha = 1$.

Now we will deal with the case $\alpha > 3/4$ and $\alpha \neq 1$. For simplicity we write

$$Q_\alpha := \frac{2^{-4\alpha-1}(3\alpha-1)}{\alpha(\alpha-1)^2} + \frac{(\alpha-1/2)B^-(1/2; 1+2\alpha, -2+2\alpha)}{2(\alpha-1)\alpha}.$$

Then, by Lemma 4.1.1 we get

$$\begin{aligned} e^{y/2}P(y) &= \frac{\alpha - \frac{1}{2}}{\alpha - 1} + \frac{e^{\frac{y}{2}}}{8(\alpha-1)\alpha}\left(\left(1 - e^{-\frac{y}{2}}\right)^{2\alpha} - 1\right) \\ &\quad - e^{-(4\alpha-3)\frac{y}{2}}\frac{B^-(1-e^{-\frac{1}{2}y}; 2\alpha, 3-4\alpha)}{4(\alpha-1)} \\ &\quad + e^{-(4\alpha-3)\frac{y}{2}}Q_\alpha + \frac{(\alpha-\frac{1}{2})^2}{4(\alpha-1)^2}e^{-\frac{y}{2}}. \end{aligned}$$

The first term is constant while the last two terms vanish as $y \to \infty$. We will therefore focus on the remaining two terms. For the first one, we have, see (4.24),

$$\frac{e^{\frac{y}{2}}}{8(\alpha-1)\alpha}\left(\left(1-e^{-\frac{y}{2}}\right)^{2\alpha} - 1\right) \sim \frac{-2\alpha}{8(\alpha-1)\alpha} = -\frac{1}{4(\alpha-1)},$$

as $y \to \infty$. Finally, writing $z = e^{-\frac{y}{2}}$ we get that

$$e^{-(4\alpha-3)\frac{y}{2}}B^-(1-e^{-\frac{1}{2}y}; 2\alpha, 3-4\alpha) = z^{4\alpha-3}B^-(1-z; 2\alpha, 3-4\alpha).$$

Therefore it follows, see Lemma B.1, that

$$\begin{aligned} \lim_{y\to\infty} -e^{-(4\alpha-3)\frac{y}{2}}\frac{B^-(1-e^{-\frac{1}{2}y}; 2\alpha, 3-4\alpha)}{4(\alpha-1)} &= \lim_{z\to 0} z^{4\alpha-3}\frac{B^-(1-z; 2\alpha, 3-4\alpha)}{4(\alpha-1)} \\ &= \frac{1}{4(\alpha-1)(4\alpha-3)}. \end{aligned}$$

We conclude that as $y \to \infty$

$$e^{y/2}P(y) \sim \frac{\alpha-\frac{1}{2}}{\alpha-1} - \frac{1}{4(\alpha-1)} - \frac{1}{4(\alpha-1)(4\alpha-3)} = \frac{1-3\alpha+2\alpha^2}{(\alpha-1)(\alpha-\frac{3}{4})} = \frac{\alpha-\frac{1}{2}}{\alpha-\frac{3}{4}},$$

which finishes the proof. □

With the asymptotic behavior of $P(y)$ we are almost ready to prove Proposition 1.5.4. First, we will prove a result that will allow us to limit the values of y, when performing the integration. For this, fix some $C > 0$ and define

$$a^{\pm}(k) = 2\log\left(\frac{k \pm C\sqrt{k\log(k)}}{\xi} \vee 1\right).$$

We will show that, as $k \to \infty$,

$$\int_0^\infty P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y = (1+o(1))\int_{a^-(k)}^{a^+(k)} P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y. \tag{4.25}$$

To establish (4.25), recall that $\mu(y) = \xi e^{\frac{y}{2}}$ and consider $\rho(y,k) = \mathbb{P}\left(\mathrm{Po}(\mu(y)) = k\right)$ as a function of y. Then, since $\mu'(y) = \mu(y)/2$, we get that

$$\frac{\partial \rho(y,k)}{\partial y} = \frac{1}{2}\left(k-\mu(y)\right)\rho(y,k),$$

which implies that $\rho(y,k)$ attains its maximum at $\mu(y) = k$. Moreover we see that the derivative is strictly positive when $\mu(y) < k$ and strictly negative when $\mu(y) > k$. Since $\mu(a^-(k)) < k$ and $\mu(a^+(k)) > k$, we conclude that $\rho(y,k)$, as a function of y, is strictly increasing on $[0, a^-(k)]$ and strictly decreasing on $[a^+(k), \infty)$. Hence, using that $P(y) \le 1$ and $e^{-\alpha a^+(k)} = O(1)$,

$$\begin{aligned}
&\int_{\mathbb{R}_+\setminus[a^-(k),a^+(k)]} P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y \\
&\quad\le \int_0^{a^-(k)} \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y + \int_{a^+(k)}^\infty \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y \\
&\quad\le \rho(a^-(k),k)\int_0^{a^-(k)} e^{-\alpha y}\,\mathrm{d}y + \rho(a^+(k),k)\int_{a^+(k)}^\infty e^{-\alpha y}\,\mathrm{d}y \\
&\quad= O(1)\left(\rho(a^-(k),k) + \rho(a^+(k),k)\right),
\end{aligned}$$

as $k \to \infty$. Next we show that

$$\rho(a^{\pm}(k),k) = O\left(k^{-(1+C^2)/2}\right).$$

Since the arguments are almost completely identical, we give the prove for $a^+(k)$.

Using Stirling's formula $k! \sim \sqrt{2\pi} k^{k+1/2} e^{-k}$ as $k \to \infty$, we write

$$\begin{aligned}
\rho(a^+(k), k) &= \frac{\mu(a^+(k))^k}{k!} e^{-\mu(a^+(k))} \\
&\sim (2\pi)^{-1/2} k^{-1/2} \left(\frac{\mu(a^+(k))}{k} \right)^k e^{-(\mu(a^+(k)-k)} \\
&= (2\pi)^{-1/2} k^{-1/2} e^{-k\left(\frac{\mu(a^+(k))}{k} - 1 - \log\left(\frac{\mu(a^+(k))}{k} \right) \right)}.
\end{aligned}$$

Since, for sufficiently large k,

$$\frac{\mu(a^+(k))}{k} = 1 + C\sqrt{\frac{\log(k)}{k}},$$

and $x - \log(1+x) \sim x^2/2$ as $x \to 0$, we get

$$\begin{aligned}
\rho(a^+(k), k) &\sim \sqrt{2\pi} k^{-1/2} e^{-k\left(C\sqrt{\frac{\log(k)}{k}} - \log\left(1 + C\sqrt{\frac{\log(k)}{k}} \right) \right)} \\
&\sim (2\pi)^{-1/2} k^{-1/2} e^{-\frac{k\left(C\sqrt{\frac{\log(k)}{k}} \right)^2}{2}} \\
&= O\left(k^{-(1+C^2)/2} \right).
\end{aligned}$$

These considerations imply that as $k \to \infty$,

$$\int_0^\infty P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y = \int_{a^-(k)}^{a^+(k)} P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y + O\left(k^{-\frac{1+C^2}{2}} \right).$$

Since $C > 0$ can be chosen arbitrarily large we conclude that

$$\int_0^\infty P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y = (1 + o(1)) \int_{a^-(k)}^{a^+(k)} P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y,$$

as $k \to \infty$. Note that this implies that if $P(y) = h(y)(1 + o(1))$, uniformly in y, as $y \to \infty$, then

$$\int_0^\infty P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y \sim \int_0^\infty h(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y, \tag{4.26}$$

as $y \to \infty$.

We now proceed with the proof of Proposition 1.5.4, which is split over the different cases for α.

Proof when $1/2 < \alpha < 3/4$ By Proposition 4.1.7 and Equation (4.26), it follows that as $k \to \infty$,

$$\gamma(k) \sim c_{\alpha,\nu} \xi^{-(4\alpha-2)} \frac{\int_0^\infty e^{-(4\alpha-2)y/2} \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}{\int_0^\infty \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y} \sim c_{\alpha,\nu} k^{-4\alpha+2}.$$

where the last line is due to (4.23) with $\beta = 2\alpha - 1$.

Proof when $\alpha = 3/4$ Similar to the previous case Proposition 4.1.7 and (4.26) imply that as $k \to \infty$

$$\gamma(k) = \frac{\int_0^\infty P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}{\int_0^\infty \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y} \sim \frac{\int_0^\infty \frac{y}{2}e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}{\int_0^\infty \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}.$$

However, the final step does not follow immediately from (4.23) because of the additional logarithmic term. To prove the result we first show that

$$\int_{a^-(k)}^{a^+(k)} P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y \sim \int_{a^-(k)}^{a^+(k)} \frac{y}{2}e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y. \tag{4.27}$$

For this we establish an upper bound for the left hand side of the form

$$\int_{a^-(k)}^{a^+(k)} \frac{y}{2}e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y \leq \frac{a^+(k)}{2}\int_{a^-(k)}^{a^+(k)} e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y$$

and similarly, a lower bound

$$\int_{a^-(k)}^{a^+(k)} \frac{y}{2}e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y \geq \frac{a^-(k)}{2}\int_{a^-(k)}^{a^+(k)} e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y$$

Now observe that as $k \to \infty$,

$$\frac{a^\pm(k)}{2} = \log\left(\frac{k \pm \sqrt{k\log(k)}}{\xi_{\alpha,\nu}}\right) \sim \log(k)$$

and therefore it follows that

$$\limsup_{k\to\infty} \frac{\int_{a^-(k)}^{a^+(k)} \frac{y}{2}e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}{\log(k)\int_{a^-(k)}^{a^+(k)} e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y} \leq 1.$$

and

$$\liminf_{k\to\infty} \frac{\int_{a^-(k)}^{a^+(k)} \frac{y}{2}e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}{\log(k)\int_{a^-(k)}^{a^+(k)} e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y} \geq 1.$$

This proves (4.27).

Next we note that by (4.23) with $\beta = 1/2$ we have

$$\frac{\int_0^\infty e^{-y/2}\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y}{\int_0^\infty \rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y} \sim \xi k^{-1}.$$

Therefore, since by (4.25),

$$\int_0^\infty P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y \sim \int_{a^-(k)}^{a^+(k)} P(y)\rho(y,k)\alpha e^{-\alpha y}\,\mathrm{d}y$$

it follows from (4.27) that as $k \to \infty$,

$$\gamma(k) \sim \frac{\int_0^\infty \frac{y}{2} e^{-y/2} \rho(y,k) \alpha e^{-\alpha y} \, \mathrm{d}y}{\int_0^\infty \rho(y,k) \alpha e^{-\alpha y} \, \mathrm{d}y}$$
$$\sim \log(k) \frac{\int_0^\infty e^{-y/2} \rho(y,k) \alpha e^{-\alpha y} \, \mathrm{d}y}{\int_0^\infty \rho(y,k) \alpha e^{-\alpha y} \, \mathrm{d}y} \sim \xi \log(k) k^{-1} = \frac{6\nu}{\pi} \log(k) k^{-1},$$

when $\alpha = 3/4$.

Proof when $\alpha > 3/4$ Again, by Proposition 4.1.7, equation (4.26) and (4.23) with $\beta = 1/2$, it follows that as $k \to \infty$,

$$\gamma(k) \sim \frac{\alpha - \frac{1}{2}}{\alpha - \frac{3}{4}} \frac{\int_0^\infty e^{-y/2} \rho(y,k) \alpha e^{-\alpha y} \, \mathrm{d}y}{\int_0^\infty \rho(y,k) \alpha e^{-\alpha y} \, \mathrm{d}y} \sim \frac{\alpha - \frac{1}{2}}{\alpha - \frac{3}{4}} \xi k^{-1} = \frac{8\alpha\nu}{\pi(4\alpha - 3)} k^{-1}.$$

4.2 Convergence of clustering coefficient and function for fixed k

We will first derive Theorem 1.5.2. It will turn out that Theorem 1.5.1 has a quick derivation assuming Theorem 1.5.2.

4.2.1 Clustering function for fixed k, proving Theorem 1.5.2

We will now show that the clustering function of the KPKVB model $c(k; G_n) \xrightarrow{\mathbb{P}} \gamma(k)$ for a fixed k. The key ideas are: First of all, the coupling of the Poissonized KPKVB model with the box model is guaranteed to be exact (in the sense that it also preserves edges) for all vertices up to height $R/4$. Secondly, when computing the expected value of the clustering function $c(k; G)$ in the subgraph of the box model induced by all vertices up to height $R/4$ using the Campbell-Mecke formula we obtain integrals that are very similar to the expressions we found earlier for $\gamma(k)$.

We will repeatedly rely on the following observation.

Lemma 4.2.1. *Let $k \geq 2$ and let G, H be graphs such that G is an induced subgraph of H, or vice versa. Then*

$$|c(k; G) - c(k; H)| \leq \frac{6|E(G)\Delta E(H)|}{N_G(k) - 2|E(G)\Delta E(H)|},$$

provided that $N_G(k) > 2|E(G)\Delta E(H)|$, where $N_G(k)$ is the number of vertices with degree k in G.

Proof. We observe that

$$
\begin{aligned}
&|c(k;G) - c(k;H)| \\
&= \left| \sum_{\substack{v\in V(G),\\ \deg_G(v)=k}} \frac{c_G(v)}{N_G(k)} - \sum_{\substack{v\in V(H),\\ \deg_H(v)=k}} \frac{c_H(v)}{N_H(k)} \right| \\
&= \left| \frac{1}{N_G(k)} \left(\sum_{\substack{v\in V(G)\setminus V(H),\\ \deg_G(v)=k}} c_G(v) + \sum_{\substack{v\in V(G)\cap V(H),\\ \deg_G(v)=k,\\ \deg_H(v)\neq k}} c_G(v) \right) \right. \\
&\qquad - \frac{1}{N_H(k)} \left(\sum_{\substack{v\in V(H)\setminus V(G),\\ \deg_H(v)=k}} c_H(v) + \sum_{\substack{v\in V(G)\cap V(H),\\ \deg_G(v)\neq k,\\ \deg_H(v)=k}} c_H(v) \right) \\
&\qquad \left. + \left(\frac{1}{N_G(k)} - \frac{1}{N_H(k)} \right) \cdot \left(\sum_{\substack{v\in V(G)\cap V(H),\\ \deg_G(v)=\deg_H(v)=k}} c_G(v) \right) \right| \\
&\leq \frac{2|E(G)\Delta E(H)|}{N_G(k)} + \frac{2|E(G)\Delta E(H)|}{N_H(k)} + \frac{|N_H(k) - N_G(k)|}{N_G(k)\cdot N_H(k)} \cdot N_H(k) \\
&= \frac{2|E(G)\Delta E(H)|}{N_G(k)} + \frac{2|E(G)\Delta E(H)|}{N_H(k)} + \frac{|N_H(k) - N_G(k)|}{N_G(k)} \\
&\leq \frac{2|E(G)\Delta E(H)|}{N_G(k)} + \frac{2|E(G)\Delta E(H)|}{N_H(k)} + \frac{2|E(G)\Delta E(H)|}{N_G(k)} \\
&\leq \frac{6|E(G)\Delta E(H)|}{N_G(k) - 2|E(G)\Delta E(H)|}.
\end{aligned}
$$

(In the second line we use that $\deg_G(v) = \deg_H(v)$ in fact implies that $c_G(v) = c_H(v)$ since one of G, H is an induced subgraph of the other. In the third line we use that the clustering coefficients $c_G(v), c_H(v)$ take values in $[0, 1]$, and if either $\deg_G(v) \neq \deg_H(v)$ or $v \in V(G)\Delta V(H)$ and v has degree K in whichever of G, H it belongs to then at least one edge of $E(G)\Delta E(H)$ is incident with v, and that every edge in $E(G)\Delta E(H)$ only affects the status of its two incident vertices. For the fifth line we used that $|N_G(k) - N_H(k)| \leq |\{v \in V(G) : \deg_G(v) = k\}\Delta\{v \in$

$V(H) : \deg_H(v) = k\}| \leq 2|E(G)\Delta E(H)|$ for similar reasons. In the last line we used that $N_H(k) \geq N_G(k) - |N_G(k) - N_H(k)| \geq N_G(k) - 2|E(G)\Delta E(H)|$.) $\square$

Lemma 4.2.2. *$|E(G_n)\Delta E(G_{Po})| = o(n)$ a.a.s.*

Proof. Let us fix some $\varepsilon > 0$ and write

$$G_- := G((1-\varepsilon)n, \alpha, (1-\varepsilon)\nu), \quad G_+ := G((1+\varepsilon)n, \alpha, (1+\varepsilon)\nu).$$

(We ignore rounding issues, i.e. the issue that $(1-\varepsilon)n, (1+\varepsilon)n$ may not be integers, to avoid notational burden. We leave the straightforward details of adapting our arguments below to deal with it to the reader.)

Observe that the vertices of $G_-, G_+, G_n, G_{\mathrm{Po}}$ all live on the same hyperbolic disk, of radius $R = 2\ln(n/\nu)$. We consider the standard coupling where we have an infinite supply of i.i.d. points $u_1, u_2, \dots$ chosen according to the (α, R)-quasi uniform distribution, the vertices of $G_n = G(n; \alpha, \nu)$ are $u_1, \dots, u_n$, the vertices of G_- are $u_1, \dots, u_{(1-\varepsilon)n}$, the vertices of G_+ are $u_1, \dots, u_{(1+\varepsilon)n}$ and the vertices of G_{Po} are $u_1, \dots, u_N$ with $N \stackrel{\mathrm{d}}{=} \mathrm{Po}(n)$ independently of $u_1, u_2, \dots$.

As $N \stackrel{\mathrm{d}}{=} \mathrm{Po}(n)$, by Chebychev's inequality, we have that $|N - n| < \varepsilon n$ a.a.s. So in particular, under our coupling we have $G_- \subseteq G_n, G_{\mathrm{Po}} \subseteq G_+$ a.a.s. We now point out that, by the results of Gugelmann et al. on the average degree ([30], Theorem 2.3), we have that a.a.s.

$$|E(G_-)| = (1-\varepsilon)^2 \cdot \frac{4\nu\alpha^2}{\pi(2\alpha-1)^2} \cdot n + o(n), \quad |E(G_+)| = (1+\varepsilon)^2 \cdot \frac{4\nu\alpha^2}{\pi(2\alpha-1)^2} \cdot n + o(n).$$

It follows that

$$|E(G_n)\Delta E(G_{\mathrm{Po}})| \leq |E(G_+) \setminus E(G_-)| = \varepsilon \cdot \frac{16\nu\alpha^2}{\pi(2\alpha-1)^2} \cdot n + o(n) \text{ a.a.s.}$$

This holds for every fixed $\varepsilon > 0$. Sending $\varepsilon \searrow 0$, concludes the proof of the lemma. $\square$

Next, let us recall that by the results of Gugelmann et al. on the degree distribution ([30], Theorem 2.2) we have that

$$\frac{N_{G_n}(k)}{n} \xrightarrow[n\to\infty]{\mathbb{P}} p(k), \tag{4.28}$$

for every fixed k. In particular $N_{G_n}(k) = \Omega(n)$ a.a.s. Combining this with lemmas 4.2.1 and 4.2.2 we obtain:

Corollary 4.2.3. *For every fixed $k \geq 2$, we have*

$$c(k; G_n) - c(k; G_{Po}) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \text{and} \quad \frac{N_{G_{Po}}(k)}{n} \xrightarrow[n\to\infty]{\mathbb{P}} p(k).$$

(For the second statement we use that $N_{G_n}(k) - 2|E(G_n)\Delta E(G_{\mathrm{Po}})| \le N_{G_{\mathrm{Po}}}(k) \le N_{G_n}(k) + 2|E(G_n)\Delta E(G_{\mathrm{Po}})|$.)

In the remainder of this section, we will denote by $G_{box,H}$ the subgraph of G_{box} induced by all vertices $(x,y) \in \mathcal{V}_{\mathrm{box}} = \mathcal{P} \cap \mathcal{R}$ of height at most $R/4$.

Lemma 4.2.4. *Under the coupling provided by Lemma 1.6.1, a.a.s., $G_{box,H}$ is an induced subgraph of G_{Po} and $|E(G_{Po}) \setminus E(G_{box,H})| = o(n)$.*

Proof. Recall that under the coupling of Lemma 1.6.1, we can view G_{box} and G_{Po} as having the same vertex set $\mathcal{V}_{\mathrm{box}} = \mathcal{P} \cap \mathcal{R}$; and two points $p = (x,y), p' = (x',y') \in \mathcal{V}_{\mathrm{box}}$ are joined by an edge in G_{box} if $|x - x'|_{\pi e^{R/2}} \le e^{(y+y')/2}$, while p, p' are joined by an edge in G_{Po} if either $y + y' \ge R$ or $y + y' < R$ and $|x - x'|_{\pi e^{R/2}} \le \Phi(y, y')$ with Φ as provided by (1.8). It follows immediately from Lemma 1.6.4 that $G_{box,H}$ is an induced subgraph of G_{Po}, a.a.s., as claimed.

Fix $\varepsilon > 0$, and let X denote the number points of $\mathcal{V}_{\mathrm{box}}$ with y-coordinate $\ge (1-\varepsilon)R$. Then X is a Poisson random variable with expectation

$$
\begin{aligned}
\mathbb{E}[X] &= \mu\left((-\frac{\pi}{2}e^{R/2}, \frac{\pi}{2}e^{R/2}] \times [(1-\varepsilon)R, R]\right) \\
&= \int_{-\frac{\pi}{2}e^{R/2}}^{\frac{\pi}{2}e^{R/2}} \int_{(1-\varepsilon)R}^{R} \left(\frac{\nu\alpha}{\pi}\right) e^{-\alpha y}\, dy\, dx \\
&= O(e^{R/2-(1-\varepsilon)\alpha R}) = o(1),
\end{aligned}
$$

the last equality holding provided ε was chosen sufficiently small (using that $\alpha > 1/2$). We conclude that, a.a.s., there are no vertices of height $\ge (1-\varepsilon)R$.

Let Y denote the number of pairs of vertices $p = (x,y), p' = (x',y') \in \mathcal{V}_{\mathrm{box}}$ with $y + y' \ge R$. Then, by the Campbell-Mecke formula

$$
\begin{aligned}
\mathbb{E}[Y] &= \int_{\mathcal{R}}\int_{\mathcal{R}} \mathbb{1}_{\{y+y' \ge R\}} \mu(dp')\mu(dp) \\
&= \int_{-\frac{\pi}{2}e^{R/2}}^{\frac{\pi}{2}e^{R/2}} \int_0^R \int_{-\frac{\pi}{2}e^{R/2}}^{\frac{\pi}{2}e^{R/2}} \int_{R-y}^{R} \left(\frac{\nu\alpha}{\pi}\right)^2 e^{-\alpha(y+y')} dy' dx' dy dx \\
&= O(Re^{(1-\alpha)R}) = o(n),
\end{aligned}
$$

the last equality holding because $\alpha > 1/2$ and $n = \nu e^{R/2}$. In particular, by Markov's inequality, $Y = o(n)$ a.a.s.

Now let Z denote the number of pairs of vertices $p = (x,y), p' = (x',y')$ for which $y + y' < R, y < (1-\varepsilon)R, R/4 \le y' < (1-\varepsilon)R$ and $|x - x'|_{\pi e^{R/2}} < \Phi(y, y')$. By Lemma 1.6.2 we have that $\Phi(y, y') = O(e^{(y+y')/2})$ for all such y, y'. By Campbell-Mecke we can write

$$
\begin{aligned}
\mathbb{E}[Z] &= \int_{-\frac{\pi}{2}e^{R/2}}^{\frac{\pi}{2}e^{R/2}} \int_0^{(1-\varepsilon)R} \int_{-\frac{\pi}{2}e^{R/2}}^{\frac{\pi}{2}e^{R/2}} \int_{R/4}^{(1-\varepsilon)R} \mathbb{1}_{\{|x-x'|_{\pi e^{R/2}} < \Phi(y,y'), y+y'<R\}} \\
&\qquad\qquad \times \left(\frac{\nu\alpha}{\pi}\right)^2 e^{-\alpha(y+y')} dy' dx dy dx' \\
&= O\left(e^{R/2} \int_0^{(1-\varepsilon)R} \int_{R/4}^{(1-\varepsilon)R} e^{(1/2-\alpha)(y+y')} dy' dy\right) \\
&= O\left(e^{R/2+(1/2-\alpha)R/4}\right) = o(n).
\end{aligned}
$$

Hence also $Z = o(n)$ a.a.s. by Markov's inequality.

This concludes the proof as we have now shown that under the stated coupling, a.a.s., $G_{box,H}$ and G_{Po} differ by only $o(n)$ edges. □

Analogously to Corollary 4.2.3 we obtain:

Corollary 4.2.5. *For every fixed $k \geq 2$ we have*

$$
c(k; G_{Po}) - c(k; G_{box,H}) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \text{and} \quad \frac{N_{G_{box,H}}(k)}{n} \xrightarrow[n\to\infty]{\mathbb{P}} p(k).
$$

Lemma 4.2.6. *For every fixed $k \geq 2$ we have*

$$
c(k; G_{box,H}) \xrightarrow[n\to\infty]{\mathbb{P}} \gamma(k).
$$

Proof. We write $\mathcal{R}_- := \mathcal{R} \cap (\mathbb{R} \times [0, R/4]) = (-\frac{\pi}{2}e^{R/2}, \frac{\pi}{2}e^{R/2}] \times [0, R/4]$ and set

$$
X := \sum_{v \in N_{G_{box,H}}(k)} c(v) = \sum_{v \in \mathcal{P}\cap\mathcal{R}_-} c_{G_{box,H}}(v) \cdot \mathbb{1}_{\{\deg_{G_{box,H}}(v)=k\}}.
$$

By the Campbell-Mecke formula

$$
\mathbb{E}[X] = \int_{\mathcal{R}_-} \mathbb{E}_{\mathcal{P}}\left[c_{G^z_{box,H}}(z) \cdot \mathbb{1}_{\{\deg_{G^z_{box,H}}(z)=k\}}\right] \mu(dz),
$$

where $G^z_{box,H}$ denotes the graph we get by adding z as an additional vertex to $G_{box,H}$, and adding edges between z and the other vertices as per the connection rule (for G_{box}). Spelling out the intensity measure μ, plus symmetry considerations, gives

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\frac{\pi}{2}e^{R/2}}^{\frac{\pi}{2}e^{R/2}} \int_0^{R/4} \mathbb{E}_{\mathcal{P}}\left[c_{G^{(x,y)}_{box,H}}((x,y)) \cdot \mathbb{1}_{\{\deg_{G^{(x,y)}_{box,H}}((x,y))=k\}} \right] \left(\frac{\nu\alpha}{\pi}\right) e^{-\alpha y}\, dy\, dx \\
&= n\int_0^{R/4} \mathbb{E}_{\mathcal{P}}\left[c_{G^{(0,y)}_{box,H}}((0,y)) \cdot \mathbb{1}_{\{\deg_{G^{(0,y)}_{box,H}}((0,y))=k\}} \right] \alpha e^{-\alpha y}\, dy \\
&= n \cdot \int_0^{R/4} \mathbb{E}_{\mathcal{P}}\left[c_{G^{(0,y_0)}_{box,H}}((0,y_0)) \middle| \deg_{G^{(0,y_0)}_{box,H}}((0,y_0)) = k \right] \cdot \\
&\qquad \mathbb{P}\left[\deg_{G^{(0,y_0)}_{box,H}}((0,y_0)) = k \right] \alpha e^{-\alpha y_0}\, dy_0.
\end{aligned}$$

The random variable $\deg_{G^{(0,y_0)}_{box,H}}((0,y_0))$ follows a Poisson distribution with expectation

$$\begin{aligned}
\mathbb{E}\left[\deg_{G^{(0,y_0)}_{box,H}}((0,y_0)) \right] &= \mu(\mathcal{B}_\infty((0,y_0)) \cap \mathcal{R}_-) \\
&= \int_0^{R/4} \int_{-e^{(y+y_0)/2}}^{e^{(y+y_0)/2}} \left(\frac{\nu\alpha}{\pi}\right) e^{-\alpha y}\, dx\, dy \\
&= \xi e^{y_0/2} \cdot (1 - e^{(1/2-\alpha)R/4}).
\end{aligned}$$

Hence, for every fixed y_0 and k, we have that

$$\mathbb{P}\left[\deg_{G^{(0,y_0)}_{box,H}}((0,y_0)) = k \right] \xrightarrow[n\to\infty]{} \mathbb{P}(\mathrm{Po}(\xi e^{y_0/2}) = k) = \rho(y_0,k).$$

Next we remark that, analogously to the argument given in the beginning of Section 4.1.2, we have

$$\mathbb{E}\left[c_{G^{(0,y_0)}_{box,H}}((0,y_0)) \middle| \deg_{G^{(0,y_0)}_{box,H}}((0,y_0)) = k \right] = \mathbb{P}(w_1 \in \mathcal{B}_\infty(w_2)) =: P_n(y_0),$$

with $w_1 = (x_1,y_1), w_2 = (x_2,y_2)$ chosen independently from $\mathcal{B}_\infty((0,y_0)) \cap \mathcal{R}_-$ according to the probability measure we get by renormalizing μ, i.e. with pdf $f_\mu \cdot \mathbb{1}_{\mathcal{B}_\infty((0,y_0))\cap\mathcal{R}_-}/\mu(\mathcal{B}_\infty((0,y_0)) \cap \mathcal{R}_-)$. By considerations completely analogous to those following Lemma 4.1.1, the random variables y_1, y_2 both follow a truncated exponential distribution with parameter $\alpha - 1/2$ truncated at height $R/4$ (i.e. with density $\mathbb{1}_{\{y_i \le R/4\}} \cdot (\alpha - 1/2)e^{(1/2-\alpha)y_i}/(1 - e^{(1/2-\alpha)R/4})$) and, given the values of y_0, y_1, y_2, each x_i is chosen uniformly on the interval $[-e^{(y_0+y_i)/2}, e^{(y_0+y_i)/2}]$. In particular

$$P_n(y_0) = \left(\frac{\alpha - 1/2}{1 - e^{(1/2-\alpha)R/4}}\right)^2 \int_0^{R/4} \int_0^{R/4} P(y_0, y_1, y_2) e^{(1/2-\alpha)(y_1+y_2)}\, dy_1\, dy_2,$$

with $P(y_0, y_1, y_2)$ as defined in the paragraph following Lemma 4.1.1. (That is, $P(y_0, y_1, y_2)$ is the probability that $|x_1 - x_2| < e^{(y_1+y_2)/2}$, where x_1, x_2 are independent with x_i uniform on the interval $[-e^{(y_0+y_i)/2}, e^{(y_0+y_i)/2}]$). It follows that, for any fixed y_0, we have

$$P_n(y_0) \xrightarrow[n\to\infty]{} (\alpha - 1/2)^2 \int_0^{\infty} \int_0^{\infty} P(y_0, y_1, y_2) e^{(1/2-\alpha)(y_1+y_2)}\, dy_1\, dy_2 = P(y_0).$$

(Applying monotone convergence to justify the convergence of the integral as $n \to \infty$.)

Since (expected) clustering coefficients and probabilities are between zero and one and $\alpha e^{\alpha y_0}$ is integrable, we can now apply the dominated convergence theorem to obtain that

$$\frac{\mathbb{E}X}{n} \xrightarrow[n\to\infty]{} \int_0^{\infty} P(y_0)\rho(y_0, k)\alpha e^{-\alpha y_0}\, dy_0 = p(k) \cdot \gamma(k). \tag{4.29}$$

(Applying (4.6) for the last equality.)

Next, we turn attention to $X(X-1) = \sum_{u \neq v \in N_{G_{box,H}}(k)} c(v)c(u)$. Another application of Campbell-Mecke shows that

$$\begin{aligned}&\mathbb{E}[X(X-1)]\\ &= \int_{\mathcal{R}_-} \int_{\mathcal{R}_-} \mathbb{E}\left[c_{G^{z,z'}_{box,H}}(z) c_{G^{z,z'}_{box,H}}(z') \cdot \mathbb{1}_{\{\deg_{G^{z,z'}_{box,H}}(z) = \deg_{G^{z,z'}_{box,H}}(z') = k\}}\right] \mu(dz)\mu(dz'),\end{aligned}$$

with $G^{z,z'}_{box,H}$ denoting the graph we get by adding z, z' as additional vertices to $G_{box,H}$. Now note that if $z = (x, y)$ and $z' = (x', y')$ satisfy $|x - x'|_{\pi e^{R/2}} > 2e^{R/4}$ then the neighbourhoods of z, z' are determined by the points of the Poisson process $\mathcal{P}$ in disjoint areas of the plane. This implies that, provided $|x - x'|_{\pi e^{R/2}} > 2e^{R/4}$:

$$\begin{aligned}&\mathbb{E}\left[c_{G^{z,z'}_{box,H}}(z) c_{G^{z,z'}_{box,H}}(z') \cdot \mathbb{1}_{\{\deg_{G^{z,z'}_{box,H}}(z) = \deg_{G^{z,z'}_{box,H}}(z') = k\}}\right]\\ &= \\ &\mathbb{E}\left[c_{G^{z}_{box,H}}(z) \cdot \mathbb{1}_{\{\deg_{G^{z}_{box,H}}(z) = k\}}\right] \cdot \mathbb{E}\left[c_{G^{z'}_{box,H}}(z') \cdot \mathbb{1}_{\{\deg_{G^{z'}_{box,H}}(z') = k\}}\right].\end{aligned} \tag{4.30}$$

On the other hand, the LHS of (4.30) is always between zero and one, also if $|x - x'|_{\pi e^{R/2}} \leq 2e^{R/4}$. We conclude that

$$\mathbb{E}[X(X-1)] \leq \int_{\mathcal{R}_-} \int_{\mathcal{R}_-} \mathbb{E}\left[c_{G^{z}_{box,H}}(z) \cdot \mathbb{1}_{\{\deg_{G^{z}_{box,H}}(z) = k\}}\right]$$

$$
\begin{aligned}
&\times \mathbb{E}\left[c_{G^{z'}_{box,H}}(z')\cdot \mathbb{1}_{\{\deg_{G^{z'}_{box,H}}(z')=k\}}\right]\mu(dz)\mu(dz') \\
&+\int_{\mathcal{R}_-}\int_{\mathcal{R}_-}\mathbb{1}_{\{|x-x'|\le 2e^{R/4}\}}\mu(dz)\mu(dz') \\
&=\left(\int_{\mathcal{R}_-}\mathbb{E}\left[c_{G^{z}_{box,H}}(z)\cdot \mathbb{1}_{\{\deg_{G^{z}_{box,H}}(z)=k\}}\right]\mu(dz)\right)^2+O(e^{3R/4}) \\
&=(\mathbb{E}[X])^2+O(n^{3/2}).
\end{aligned}
$$

Combining this with (4.29), it follows that $\mathrm{Var}(X)=\mathbb{E}[X(X-1)]+\mathbb{E}[X]-(\mathbb{E}[X])^2=o\left((\mathbb{E}[X])^2\right)$. By Chebychev's inequality, we therefore have

$$X=n\cdot\gamma(k)\cdot p(k)+o(n) \text{ a.a.s.}$$

In combination with Corollary 4.2.5 (second limit) we can conclude that

$$c(k;G_{box,H})=\frac{X}{N_{G_{box,H}(k)}}\xrightarrow[n\to\infty]{\mathbb{P}}\gamma(k),$$

as desired. □

Proof of Theorem 1.5.2: For completeness, we point out that Theorem 1.5.2 follows immediately from Corollaries 4.2.3, 4.2.5 and Lemma 4.2.6. ■

4.2.2 Overall clustering coefficient, proving Theorem 1.5.1

Proof of Theorem 1.5.1: Recall in Section 4.1, we *defined* $p(k):=\mathbb{P}(D=k),\gamma:=\mathbb{E}C,\gamma(k):=\mathbb{E}(C|D=k)$ with D the degree and C the clustering coefficient of the "typical point" in the infinite limit model G_∞. We can write

$$\gamma=\mathbb{E}C=\sum_{k\ge 2}\mathbb{E}\left(C|D=k\right)\mathbb{P}(D=k)=\sum_{k\ge 2}\gamma(k)\cdot p(k).$$

For the KPKVB random graph, or any graph for that matter, we have the similar relation

$$c(G_n)=\sum_{k\ge 2}c(k;G_n)\cdot(N_{G_n}(k)/n).$$

By Theorem 1.5.2 and (4.28) we have, for any fixed $k\ge 2$:

$$c(G_n)\ge\sum_{k=2}^{K}c(k;G_n)\cdot(N_{G_n}(k)/n)\xrightarrow[n\to\infty]{\mathbb{P}}\sum_{k=2}^{K}\gamma(k)\cdot p(k), \tag{4.31}$$

where Slutsky's theorem justifies the convergence in probability. On the other hand we have

$$c(G_n) \le \sum_{k=2}^{K} c(k;G_n)\cdot(N_{G_n}(k)/n) + \sum_{k>K}(N_{G_n}(k)/n) \xrightarrow[n\to\infty]{\mathbb{P}} \sum_{k=2}^{K}\gamma(k)\cdot p(k) + \sum_{k>K} p(k), \tag{4.32}$$

where the convergence in probability can be justified using Slutsky's theorem together with the fact that $\sum_{k=0}^{\infty} p(k) = 1$ (one convenient way to convince oneself that this is true, is to note that D, the degree of the typical point, is a.s. finite). In more detail,

$$\sum_{k>K}(N_{G_n}(k)/n) = 1 - \sum_{k=0}^{K}(N_{G_n}(k)/n) \xrightarrow[n\to\infty]{\mathbb{P}} 1 - \sum_{k=0}^{K} p(k) = \sum_{k>K} p(k).$$

The result follows from (4.32) and (4.31), by sending $K \to \infty$. ■

4.3 Overview of the proof strategy for $k \to \infty$

The proof of Theorem 1.5.3 follows the same strategy as outlined in Figure 4.2. The fact that $k = k_n \to \infty$ as $n \to \infty$, introduces significant technical challenges, especially for k_n close to the maximum scaling of $n^{\frac{1}{2\alpha+1}}$. For example, the coupling between G_{Po} and G_{box} we use becomes less exact so that we can no longer use Lemma 1.6.4 to conclude that triangle counts in G_{Po} and G_{box} are asymptotically equivalent. Furthermore, it will require a great deal of care to bound all error terms we encounter.

In this section we explain the challenges associated with each step and give a detailed overview of the structure for the proof of Theorem 1.5.3 using intermediate results for each of the steps. To this end we define the scaling function

$$s(k) = \begin{cases} k^{-(4\alpha-2)}, & \text{if } \frac{1}{2} < \alpha < \frac{3}{4}, \\ \log(k)k^{-1}, & \text{if } \alpha = \frac{3}{4}, \\ k^{-1}, & \text{if } \alpha > \frac{3}{4}, \end{cases} \tag{4.33}$$

so that $\gamma(k) = \Theta(s(k))$ as $k \to \infty$. We will end this section with the proof of Theorem 1.5.3, based on the intermediate results.

Remark 4.3.1 (Diverging k_n)**.** Throughout the remainder of this chapter, unless stated otherwise, $(k_n)_{n\ge1}$ will always denote a sequence of non-negative integers satisfying $k_n \to \infty$ and $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$, as $n \to \infty$.

We start with introducing a slightly adjusted version of the local clustering function, which will be convenient for computations later,

$$c^*(k;G) = \frac{1}{\mathbb{E}[N(k)]} \sum_{\substack{v\in V(G) \\ \deg(v)=k}} c(v). \tag{4.34}$$

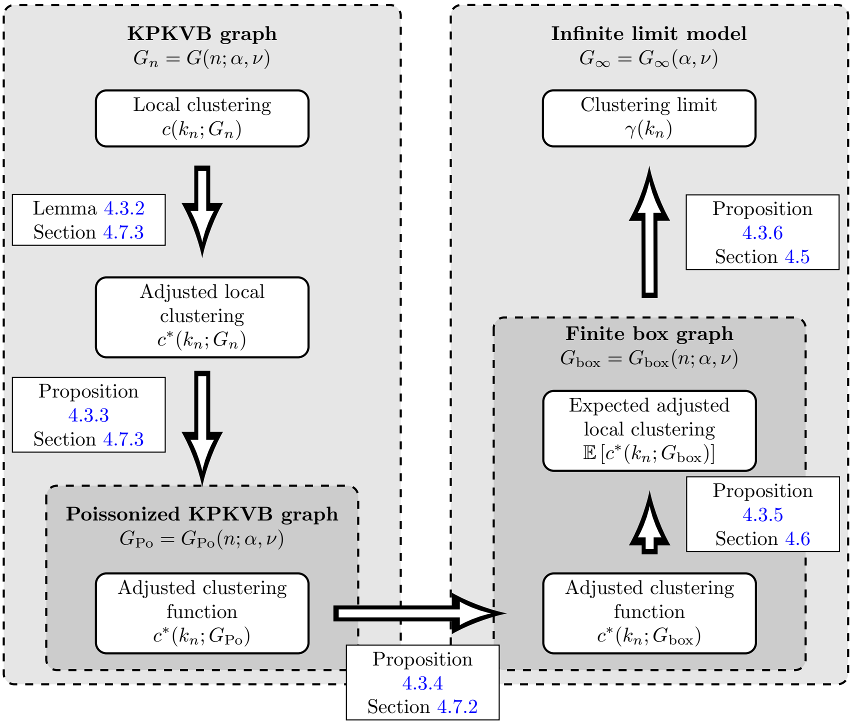

Figure 4.2: Overview of the proof strategy for Theorem 1.5.2. The left column denote the models in which the true hyperbolic balls are used while the right column contains the models that use an approximation of these. The most important part is the transition between these two settings which is accomplished by Proposition 4.3.4.

Notice that the only difference between $c(k;G)$ and $c^*(k;G)$ is that we replace $N(k)$ by its expectation $\mathbb{E}[N(k)]$. The advantage is that now, the only randomness is in the formation of triangles. In addition, note that since $\mathbb{E}[N(k)] > 0$ a case distinction for $N(k)$ is no longer needed for $c^*(k;G)$. It is however still relevant since we are eventually interested in $c(k;G)$. Following the notational convention, throughout the remainder of this chapter, we write $c^*(k;G_{\mathrm{Po}})$ and $c^*(k;G_{\mathrm{box}})$ to denote the adjusted local clustering function in G_{Po} and G_{box}, respectively.

Figure 4.2 shows a schematic overview of the proof of Theorem 1.5.3 based on the different propositions described below, plus the sections in which these propositions are proved. Observe that the order in which the intermediate results are proved is reversed with respect to the natural order of reasoning. This does not create any circular logic, since each intermediate result is independent of all others. We choose this order because results proved in the later stages are helpful to deal with error terms coming up in proofs at earlier stages and hence help streamline those proofs. Below, we briefly describe each of the intermediate steps leading up to the proof of Theorem 1.5.3. We start with an observation on the dependence between the degree of a point $p = (x, y)$ and its height y.

4.3.1 Restricting the height for vertices with degree k_n

In the proof of Proposition 1.5.4 we have used a result that allowed us to restrict integration over y to the interval $[a^-(k), a^+(k)]$, with

$$a^{\pm}(k) = 2\log\left(\frac{k \pm C\sqrt{k\log(k)}}{\xi} \vee 1\right).$$

The reason for this was that the integrand included the function

$$\rho(y, k) = \mathbb{P}\left(\mathrm{Po}(\mu(y)) = k\right),$$

where $\mathrm{Po}(\lambda)$ denotes a Poisson random variable with expectation λ and Poisson random variables are well concentrated around their expectation, i.e. their probability is largest around heights y and degrees k for which $\mu(y) \approx k$. Since $\mu(y) = \mu(\mathcal{B}_\infty(y)) = \xi e^{y/2}$, this implies that integration with respect to $\rho(y, k)$ is concentrated around $y \approx 2\log(k/\xi)$. In the remainder of this chapter, we will often encounter integrands involving the function $\mathbb{P}\left(\mathrm{Po}(\mu_n(y)) = k_n\right) =: \hat{\rho}_n(y, k_n)$, for some $\mu_n(y)$ which is asymptotically equivalent to $\mu(y)$. In these cases we also want to be able to restrict our integration around those heights y for which $y \approx 2\log(k_n/\xi)$. Such results are established in Section 4.4. Moreover, we prove that if μ_n corresponds to either the intensity measure $\mu\left(\mathcal{B}(y)\right)$ in the Poissonized KPKVB model G_{Po} or the intensity measure $\mu\left(\mathcal{B}_{\mathrm{box}}(y)\right)$ in the box model G_{box}, then for a certain class of functions h,

$$\int_0^\infty h(y)\hat{\rho}_n(y, k_n)\alpha e^{-\alpha y}\,\mathrm{d}y = (1 + o(1))\int_0^\infty h(y)\rho(y, k_n)\alpha e^{-\alpha y}\,\mathrm{d}y,$$

i.e. we may replace $\hat{\rho}_n(y, k_n)$ in the integrand by $\rho(y, k_n)$.

4.3.2 Adjusted clustering and the Poissonized KPKVB model

Recall that the first step for the fixed k case was to show that the transition from the KPKVB graph $G_n = G(n; \alpha, \nu)$ to the Poissonized version G_{Po} did not influence clustering. Here we first make a transition from the local clustering function $c(k_n; G_n)$ to the adjusted version $c^*(k_n; G_n)$, for the proof see Section 4.7.3.

Lemma 4.3.2. *As $n \to \infty$,*

$$\left|c(k_n; G_n) - c^*(k_n; G_n)\right| = o_{\mathbb{P}}\left(s(k_n)\right).$$

We then establish that the adjusted local clustering function for KPKVB graphs G_n behaves similarly to that in G_{Po}. The proof, found in Section 4.7.3, is based on a standard coupling between a Binomial Point Process and Poisson Point Process.

Proposition 4.3.3. As $n \to \infty$,

$$\mathbb{E}\left[|c^*(k_n; G_n) - c^*(k_n; G_{\text{Po}})|\right] = o\left(s(k_n)\right).$$

4.3.3 Coupling of local clustering between G_{Po} and G_{box}

The next step is to show that the adjusted clustering is preserved under the coupling described in Section 1.6.4. The proof can be found in Section 4.7.2. This step is one of the key technical challenges we face in proving Theorem 1.5.3.

To understand why, recall that the degree k of a vertex is related to its height y, roughly speaking, by $k \approx \xi e^{y/2}$. Therefore, when k is fixed we have that the heights of vertices with that degree are also fixed, in particular $y < R/4$ for large enough n. In addition, the main contribution of triangles would also come from vertices with heights $y' < R/4$. This allowed us to use Lemma 1.6.4 and to conclude that the triangles present in the graph G_{Po} were exactly those present in G_{box} and therefore the local clustering function was the same in both models. When $k_n \to \infty$ this is no longer true in general. For instance, suppose that $k_n = n^{\frac{1-\varepsilon}{2\alpha+1}}$, for some small $0 < \varepsilon < 1$. Then the relation $k_n \approx \xi e^{y_n/2}$ implies that $y_n \approx \frac{2(1-\varepsilon)}{2\alpha+1} \log(n) - 2\log(\xi)$. Since $R/4 = \frac{1}{2}\log(n) - \frac{1}{2}\log(\nu)$ we get that $R/4 = o\left(y_n\right)$ for all $\alpha > (3-4\varepsilon)/2$ and hence $y_n > R/4$ for large enough n, violating the conditions of Lemma 1.6.4. However, by carefully analyzing the difference between the adjusted local clustering function in both models we can still make the same conclusion. This is summarized in the following proposition whose proof is in Section 4.7.2:

Proposition 4.3.4 (Coupling result for adjusted clustering function)**.** As $n \to \infty$,

$$\mathbb{E}\left[|c^*(k_n; G_{\text{Po}}) - c^*(k_n; G_{\text{box}})|\right] = o\left(s(k_n)\right).$$

Together, the three results described so far imply that the difference between the clustering function for a KPKVB graph and the adjusted clustering function for the finite box graph G_{box} converges to zero faster than the proposed scaling $\gamma(k_n)$ in Theorem 1.5.3. Hence, to prove this theorem it is enough to prove it for $c^*(k_n; G_{\text{box}})$.

4.3.4 From the finite box to the infinite model

To compute the limit of the adjusted clustering function $c^*(k_n; G_{\text{box}})$ we first prove in Section 4.6 that it is concentrated around its expectation $\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right]$.

Proposition 4.3.5 (Concentration for adjusted clustering function in G_{box})**.** As $n \to \infty$,

$$\mathbb{E}\left[\left|c^*(k_n; G_{\text{box}}) - \mathbb{E}\left[c^*(k_n; G_{\text{box}})\right]\right|\right] = o\left(s(k_n)\right).$$

This result represents another technical challenge we face when considering $k_n \to \infty$. For the proof, we first identify the specific range of heights that give the main contribution to the triangle count, showing that the triangles coming from vertices with heights outside this range is of smaller order. Then we prove a concentration result for the main term, by carefully analyzing the joint neighbourhoods of two vertices whose heights fall into the identified range. The full details are in Section 4.6.

Assuming this concentration result, we are left to compute the expectation $\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right]$ and show that it is asymptotically equivalent to $\gamma(k_n)$ as $n \to \infty$. To accomplish this, we move to the infinite limit model G_∞ and show that the difference between the expected value of $c^*(k; G_{\text{box}})$ and $\gamma(k_n)$ goes to zero faster than the proposed scaling in Theorem 1.5.2.

Proposition 4.3.6 (Transition to the infinite limit model)**.** As $n \to \infty$,

$$\left|\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right] - \gamma(k_n)\right| = o\left(s(k_n)\right).$$

Recall that for the finite box model the left and right boundaries of $\mathcal{R}_n$ were identified, so that the graph G_{box} contains some additional edges with respect to the induced subgraph of G_∞ on $\mathcal{R}_n$. The proof of Proposition 4.3.6 therefore relies on analyzing the number of triangles coming from these additional edges and showing that their contribution to the clustering function are of negligible order, see Section 4.5.

Remark 4.3.7 (Notations for different graphs)**.** We will use the subscripts n, Po, box and ∞ to identify properties of, respectively, the KPKVB model G_n, the Poisson version G_{Po}, the finite box model G_{box} and the infinite model G_∞. For example $N_{\text{Po}}(k)$ denotes the number of vertices with degree k in G_{Po} and $\rho_{\text{box}}(y,k) = \mathbb{P}\left(\text{Po}(\mu(\mathcal{B}_{\text{box}}(y))) = k\right)$, i.e. the degree distribution in G_{box} for a point $p = (x, y)$.

4.3.5 Proof of the main results

We are now ready to prove Theorem 1.5.3, using the results stated in the previous sections.

Proof of Theorem 1.5.3. Note that the second statement of the theorem ('In particular, ...') follows immediately from the first one.

To prove the first statement, we rewrite $c(k_n; G_n) - \gamma(k_n)$ as

$$\begin{aligned} c(k_n; G_n) - \gamma(k_n) = {} & (c(k_n; G_n) - c^*(k_n; G_n)) + (c^*(k_n; G_n) - c^*(k_n; G_{\text{Po}})) \\ & + (c^*(k_n; G_{\text{Po}}) - c^*(k_n; G_{\text{box}})) \end{aligned}$$

$$
\begin{aligned}
&+ \left(c^*(k_n; G_{\text{box}}) - \mathbb{E}[c^*(k_n; G_{\text{box}})]\right) \\
&+ \mathbb{E}[c^*(k_n; G_{\text{box}})] - \gamma(k_n).
\end{aligned}
$$

Then, we take absolute values and apply the triangle inequality. By monotonicity of expectation, we can apply it to both sides and obtain

$$
\begin{aligned}
\mathbb{E}\left[|c(k_n; G_n) - \gamma(k_n)|\right] \leq\ & \mathbb{E}\left[|c(k_n; G_n) - c^*(k_n; G_n)|\right] + \mathbb{E}\left[|c^*(k_n; G_n) - c^*(k_n; G_{\text{Po}})|\right] \\
&+ \mathbb{E}\left[|c^*(k_n; G_{\text{Po}}) - c^*(k_n; G_{\text{box}})|\right] \\
&+ \mathbb{E}\left[|c^*(k_n; G_{\text{box}}) - \mathbb{E}c^*(k_n; G_{\text{box}})|\right] \\
&+ \left|\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right] - \gamma(k_n)\right|.
\end{aligned}
$$

At this point, the lemmas and propositions presented above in this section can be applied in order to show that all summands are $o\left(\gamma(k_n)\right)$: Lemma 4.3.2 for the transition to the modified clustering function in the first term, Proposition 4.3.3 for the Poissonization in the second term, Proposition 4.3.4 for the coupling between the Poissonized KPKVB and the finite box model in the third term, Proposition 4.3.5 for the concentration in the fourth term and finally Proposition 4.3.6 for the transition to the infinite limit model.

All of this together yields that

$$
\mathbb{E}\left[|c(k_n; G_n) - \gamma(k_n)|\right] = o\left(s(k_n)\right) = o\left(\gamma(k_n)\right),
$$

which establishes the first statement of the theorem and finishes the proof. □

4.4 Concentration of heights for vertices with degree k

Here we show that if we integrate with respect to the function

$$
\hat{\rho}_n(y, k_n) = \mathbb{P}\left(\text{Po}(\mu_n(y)) = k_n\right),
$$

then we may restrict integration with respect to the *height* y to an interval on which $\mu_n(y) = \Theta\left(k_n\right)$. We will refer to such a result as a *concentration of heights* result. In addition, if $\mu_n(y)$ is equivalent to $\mu(y)$ on this interval, then we may replace $\hat{\rho}_n(y, k_n)$ in the integral by $\rho(y, k_n) = \mathbb{P}\left(\text{Po}(\mu(y)) = k_n\right)$ (the degree distribution of the typical point in G_∞).

We start with a concentration of heights result for the infinite model G_∞ (Lemma 4.4.1) and explain in Remark 4.4.5 how such a result will be used throughout the chapter. We then present a generalization of this result (Lemma 4.4.7) and use this to establish concentration of heights results for the Poissonized KPKVB G_{Po} and finite box model G_{box}.

Finally we provide a general result that allows to substitute $\hat{\rho}_n(y, k_n)$ in the integrand with $\rho(y, k_n)$ and show that this holds in particular for the degree distributions in G_{Po} and G_{box}, given by, respectively $\rho_{\text{Po}}(y, k_n) := \mathbb{P}\left(\text{Po}(\mu\left(\mathcal{B}\left(y\right)\right)) = k_n\right)$ and $\rho_{\text{box}}(y, k_n) := \mathbb{P}\left(\text{Po}(\mu\left(\mathcal{B}_{\text{box}}\left(y\right)\right)) = k_n\right)$.

4.4.1 Concentration of heights argument for the infinite model

The next lemma states that for a large class of functions $h(y)$ and $k_n \to \infty$, to compute the integral

$$\int_0^\infty \rho(y, k_n) h(y) e^{-\alpha y} \, \mathrm{d}y,$$

it is enough to consider integration over a small interval on which $e^{y/2} \approx k_n$, instead of $\mathbb{R}_+$.

Lemma 4.4.1. *Let $\alpha > \frac{1}{2}$, $\nu > 0$, $(k_n)_{n \geq 1}$ be any positive sequence such that $k_n \to \infty$ and $k_n = o(n)$ and let $\ell_n = k_n(1 + \epsilon_n)$, with $\epsilon_n \to 0$. In addition, define for any constant $C > 0$,*

$$\lambda_n^{\pm} = (\ell_n \pm C\sqrt{\ell_n \log(\ell_n)}) \wedge \xi, \quad a_n^{\pm} = 2 \log\left(\frac{\lambda_n^{\pm}}{\xi}\right).$$

Then the following holds:

1. *For any continuous function $h : \mathbb{R}_+ \to \mathbb{R}$, such that $h(y) = O\left(e^{\beta y}\right)$ as $y \to \infty$ for some $\beta < \alpha$,*

$$\int_{\mathbb{R}_+ \setminus (a_n^-, a_n^+)} \rho(y, k_n) h(y) \alpha e^{-\alpha y} \, \mathrm{d}y = O\left(k_n^{-(1+C^2)/2}\right), \tag{4.35}$$

as $n \to \infty$.

2. *If in addition $C > \sqrt{4\alpha + 1}$ and $h(a_n) \sim h(b_n)$ whenever $a_n \sim b_n$, as $n \to \infty$. Then,*

$$\int_0^\infty h(y) \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y \sim 2\alpha \xi^{2\alpha} h(2 \log(k_n/\xi)) k_n^{-(2\alpha+1)}, \tag{4.36}$$

as $n \to \infty$.

Proof.

Proof of the first statement. Recall (see proof of Proposition 1.5.4) that $\rho(y, k_n)$, as a function of y, is strictly increasing on $[0, a_n^-]$ and strictly decreasing on $[a_n^+, \infty)$. Therefore, by our assumption on $h(y)$,

$$\begin{aligned}
&\int_{\mathbb{R}_+ \setminus (a_n^-, a_n^+)} h(y) \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&\quad = O(1) \int_0^{a_n^-} e^{\beta y} \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y + O(1) \int_{a_n^+}^\infty e^{\beta y} \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&\quad = O(1) \int_0^{a_n^-} \rho(y, k_n) e^{-(\alpha - \beta) y} \, \mathrm{d}y + O(1) \int_{a_n^+}^\infty \rho(y, k_n) e^{-(\alpha - \beta) y} \, \mathrm{d}y
\end{aligned}$$

$$\leq O(1)\,\rho(a_n^-,k_n)\int_0^{a_n^-} e^{-(\alpha-\beta)y}\,\mathrm{d}y + O(1)\,\rho(a_n^+,k_n)\int_{a_n^+}^{\infty} e^{-(\alpha-\beta)y}\,\mathrm{d}y.$$

Since $\alpha-\beta>0$, we conclude that

$$\int_{\mathbb{R}_+\setminus(a_n^-,a_n^+)} h(y)\rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y = O(1)\left(\rho(a_n^-,k_n)+\rho(a_n^+,k_n)\right). \tag{4.37}$$

We shall now bound the terms $\rho(a_n^{\pm},k_n)$. We explicitly show the bound for $\rho(a_n^+,k_n)$, the computation for $\rho(a_n^-,k_n)$ is similar. Using Stirling's approximation $k!\sim\sqrt{2\pi}k^{k+1/2}e^{-k}$ as $k\to\infty$ we write

$$\begin{aligned}\rho(a_n^+,k_n) &= \frac{\mu(a_n^+)^{k_n}}{k_n!}e^{-\mu(a_n^+)}\\ &\sim (2\pi)^{-1/2}k_n^{-1/2}\left(\frac{\mu(a_n^+)}{k_n}\right)^{k_n}e^{-(\mu(a_n^+)-k_n)}\\ &= (2\pi)^{-1/2}k_n^{-1/2}e^{-k_n\left(\frac{\mu(a_n^+)}{k_n}-1-\log\left(\frac{\mu(a_n^+)}{k_n}\right)\right)}.\end{aligned}$$

Since for $\kappa_n=\sqrt{(1+\epsilon_n)k_n\log((1+\epsilon_n)k_n)}$,

$$\frac{\mu(a_n^+)}{k_n}=\frac{\lambda_n^+}{k_n}=1+\epsilon_n+C\frac{\kappa_n}{k_n}=1+\epsilon_n+C\sqrt{\frac{(1+\epsilon_n)\log((1+\epsilon_n)k_n)}{k_n}},$$

and as $x-\log(1+x)\sim x^2/2$ as $x\to0$, we get

$$\begin{aligned}\rho(a_n^+,k_n) &\leq \sqrt{2\pi}k_n^{-1/2}e^{-k_n\left(\epsilon_n+C\frac{\kappa_n}{k_n}-\log\left(1+\epsilon_n+C\frac{\kappa_n}{k_n}\right)\right)}\\ &\sim (2\pi)^{-1/2}k_n^{-1/2}e^{-\frac{k_n(\epsilon_n+C\kappa_n/k_n)^2}{2}}\\ &= O\left(k_n^{-(1+C^2)/2}\right),\end{aligned} \tag{4.38}$$

where for the last line we used that

$$-k_n\frac{(\epsilon_n+C\kappa_n/k_n)^2}{2}=-\frac{C^2}{2}\log(k_n)+\Theta(1).$$

A similar analysis as above yields

$$\rho(a_n^-,k_n)\leq\Theta(1)\,k_n^{-1/2}e^{-\frac{k_n(\epsilon_n-C\kappa_n/k_n)^2}{2}}=O\left(k_n^{-(1+C^2)/2}\right). \tag{4.39}$$

Plugging (4.39) and (4.38) into (4.37) yields the result.

Proof of the second statement. By the mean value theorem for definite integrals, there exists a $c_n \in (a_n^-, a_n^+)$ such that

$$\int_{a_n^-}^{a_n^+} h(y)\rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y = h(c_n)\int_{a_n^-}^{a_n^+} \rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y.$$

Since $\int_0^\infty \rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y = \Theta\left(k_n^{-(2\alpha+1)}\right)$, taking any $C > \sqrt{4\alpha+1}$, (4.35) implies that

$$\int_{a_n^-}^{a_n^+} \rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y = (1+o(1))\int_0^\infty \rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y,$$

from which we conclude that (see (4.4)),

$$\int_{a_n^-}^{a_n^+} \rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y = (1+o(1))2\alpha\xi^{2\alpha}k_n^{-(2\alpha+1)},$$

as $n\to\infty$. Finally, since $c_n \in (a_n^-, a_n^+)$ it follows that

$$\left|\frac{c_n}{2\log(k_n/\xi)} - 1\right| \leq 2C\sqrt{\frac{\log(k_n)}{k_n}},$$

so that $c_n \sim 2\log(k_n/\xi)$. Therefore, by assumption on h,

$$\int_{a_n^-}^{a_n^+} h(y)\rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y \sim h(c_n)2\alpha\xi^{2\alpha}k_n^{-(2\alpha+1)} \sim 2\alpha\xi^{2\alpha}h(2\log(k_n/\xi))k_n^{-(2\alpha+1)},$$

as $n\to\infty$. □

Note that we can tune the error in (4.35) by selecting an appropriately large $C > 0$, i.e. by restricting the function $h(y)$ inside the integral to an appropriate interval around $2\log(k_n/\xi)$. This makes Lemma 4.4.1 very powerful. Below we list several important corollaries.

Corollary 4.4.2. *Let $h : \mathbb{R}_+ \to \mathbb{R}$ be any continuous function such that for some $\beta < \alpha$, $h(y) = O\left(e^{\beta y}\right)$ as $y\to\infty$ and $h(a_n) \sim h(b_n)$ whenever $a_n \sim b_n$. Then for any other continuous function $g : \mathbb{R}_+ \to \mathbb{R}$, such that $g(y) \sim h(y)$ as $y \to \infty$*

$$\int_0^\infty g(y)\rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y \sim 2\alpha\xi^{2\alpha}h(2\log(k_n/\xi))k_n^{-(2\alpha+1)}, \tag{4.40}$$

as $n\to\infty$.

Proof. By assumption, the function g satisfies the conditions of the second statement of Lemma 4.4.1. Since in addition $g(2\log(k_n/\xi)) \sim h(2\log(k_n/\xi))$, the result follows. □

For any $C > 0$ we define

$$\mathcal{K}_C(k_n) = \left\{ y \in \mathbb{R}_+ : \frac{k_n - C\sqrt{k_n \log(k_n)}}{\xi} \vee 1 \leq e^{\frac{y}{2}} \leq \frac{k_n + C\sqrt{k_n \log(k_n)}}{\xi} \right\}. \tag{4.41}$$

The following corollary allows us to bound integrals of functions $h_n(y)$ by considering their maximum of $\mathcal{K}_C(k_n)$.

Corollary 4.4.3. *Let $h_n : \mathbb{R}_+ \to \mathbb{R}_+$ be a sequence of continuous functions such that for some $s \in \mathbb{R}$ and $\beta < \alpha$, as $n \to \infty$, $h_n(y) = O\left(k_n^s e^{\beta y}\right)$ and $h_n(y) = \Omega(1)$, uniformly on $0 \leq y \leq R$. Recall that $f(x, y) = \frac{\alpha\nu}{\pi} e^{-\alpha y}$ denotes the intensity function of the infinite limit model, as defined in (1.3). Then, as $n \to \infty$,*

$$\int_{\mathcal{R}} h_n(y)\rho(y, k_n) f(x, y) \, \mathrm{d}x \, \mathrm{d}y = (1 + o(1))\, n \int_{\mathcal{K}_C(k_n)} h_n(y)\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y.$$

In particular,

$$\int_{\mathcal{R}} h_n(y)\rho(y, k_n) f(x, y) \, \mathrm{d}x \, \mathrm{d}y = O(1)\, n \, k_n^{-(2\alpha+1)} \max_{y \in \mathcal{K}_C(k_n)} h_n(y),$$

as $n \to \infty$.

Proof. The second result follows immediately from the first. For the first result we note that, by the first statement of Lemma 4.4.1,

$$\begin{aligned}\int_{[0,R]\setminus(a_n^-, a_n^+)} h_n(y)\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y &\leq O\left(k_n^s\right) \int_{\mathbb{R}_+\setminus(a_n^-, a_n^+)} e^{\beta y}\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y \\ &= O\left(k_n^{s-(1+C^2)/2}\right).\end{aligned}$$

By assumption on $h_n(y)$,

$$\begin{aligned}\int_{\mathcal{K}_C(k_n)} h_n(y)\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y &= O\left(k_n^{s+2\beta}\right) \int_{\mathcal{K}_C(k_n)} \rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y \\ &= O\left(k_n^{s+2\beta-(2\alpha+1)}\right),\end{aligned}$$

and

$$\int_{\mathcal{K}_C(k_n)} h_n(y)\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y = \Omega(1) \int_{\mathcal{K}_C(k_n)} \rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y = \Omega(k_n^{-(2\alpha+1)}).$$

Hence, by taking $C > 0$ such that $(1 + C^2)/2 > \max\{2\alpha + 1 + s, 2\alpha + 1 - \beta\}$ we get that

$$\int_{[0,R]\setminus(a_n^-, a_n^+)} h_n(y)\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y = o(1) \int_{\mathcal{K}_C(k_n)} h_n(y)\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y.$$

The result then follows since

$$\int_{\mathcal{R}} h_n(y)\rho(y,k_n)f(x,y)\,\mathrm{d}x\,\mathrm{d}y = n\int_0^R h_n(y)\rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y.$$

□

For functions $h_n(y) = k_n^s h(y)$ we obtain an asymptotically equivalent expression for the associated integral:

Corollary 4.4.4. *Let $h : \mathbb{R}_+ \to \mathbb{R}$ be a continuous function which satisfies the conditions of Lemma 4.4.1 and let h_n be a sequence of functions such that, for some $s \in \mathbb{R}$, as $n \to \infty$, $h_n(y) = \Omega(1)$ and $h_n(y) = O\left(k_n^s\right) h(y)\rho(y,k_n)$, uniformly on $0 \le y \le R$. Then,*

$$\int_{\mathcal{R}} h_n(y)\rho(y,k_n)f(x,y)\,\mathrm{d}x\,\mathrm{d}y \sim 2\alpha\xi^{2\alpha}\, n\, h_n(2\log(k_n/\xi))k_n^{-(2\alpha+1)}, \tag{4.42}$$

as $n \to \infty$.

Proof. The result immediately follows by first applying Corollary 4.4.3 and then using the second statement from Lemma 4.4.1. □

Remark 4.4.5 (Concentration of heights argument)**.** All the above corollaries use the same reasoning, namely that when the integrand contains $h_n(y)\rho(y,k_n)$, for some "nice" functions $h_n(y)$, then the main contribution is determined by the integration over $\mathcal{K}_C(k_n)$. This implies, for instance, that we only need to carefully analyze the functions $h_n(y)$ on $\mathcal{K}_C(k_n)$, while for a certain class of functions we can even simply replace it with $h_n(2\log(k_n/\xi))$. We will refer collectively to any of these arguments as a *concentration of heights argument.*

Remark 4.4.6 (Proof of Proposition 1.5.4 revisited)**.** Note that due to Proposition 4.1.7, the function $P(y)$ from Section 4.1 satisfies all the necessary conditions in Corollary 4.4.2 with

$$h(y) := \begin{cases} e^{-\frac{y}{2}(4\alpha-2)}c_\alpha\xi^{4\alpha-2} & \text{if } \frac{1}{2} < \alpha < \frac{3}{4}, \\ \frac{y}{2}e^{-\frac{y}{2}} & \text{if } \alpha = \frac{3}{4}, \\ e^{-\frac{y}{2}}\frac{\alpha-\frac{1}{2}}{\alpha-\frac{3}{4}} & \text{if } \alpha > \frac{3}{4}. \end{cases}$$

Hence, Proposition 1.5.4 directly follows from Proposition 4.1.7 and a concentration of heights argument (as in Corollary 4.4.2).

4.4.2 A more general concentration of heights argument

Although powerful, the current versions of the concentration of heights arguments are only valid for the function $\rho(y,k_n) := \mathbb{P}\left(\mathrm{Po}\left(\mu\left(\mathcal{B}_\infty\left(y\right)\right)\right) = k_n\right)$, which uses the neighbourhoods in the infinite model G_∞. Since we will also be working in the Poissonized KPKVB model G_{Po} and the finite box model G_{box}, we would like to use concentration of heights arguments for the degree distribution function in these models. To be more precise, let us define

$$\rho_{\mathrm{Po}}(y,k) = \mathbb{P}\left(\mathrm{Po}(\mu(\mathcal{B}\left(y\right))) = k\right)$$

and

$$\rho_{\mathrm{box}}(y,k) = \mathbb{P}\left(\mathrm{Po}(\mu(\mathcal{B}_{\mathrm{box}}\left(y\right))) = k\right)$$

Then we want Lemma 4.4.1 to remain true if we replace $\rho(y,k_n)$ with either the function $\rho_{\mathrm{Po}}(y,k_n)$ or $\rho_{\mathrm{box}}(y,k_n)$. To establish this result we first prove a more general version of Lemma 4.4.1:

Lemma 4.4.7. *Let $\alpha > \frac{1}{2}, \nu > 0$ and $0 < \varepsilon < 1$. Furthermore, let $k_n \to \infty$ such that $k_n = O\left(n^{1-\varepsilon}\right)$, $\ell_n = (1+\epsilon_n)k_n$, with $\epsilon_n \to 0$ and define*

$$\lambda_n^{\pm} = (\ell_n \pm C\sqrt{\ell_n \log(\ell_n)}) \wedge \xi, \quad \text{and} \quad a_n^{\pm} = 2\log\left(\frac{\lambda_n^{\pm}}{\xi}\right),$$

for some $C > 0$. Finally, let $\hat{\rho}_n(y,k) = \mathbb{P}\left(\mathrm{Po}(\hat{\mu}_n(y)) = k\right)$, where $\hat{\mu}_n(y)$ is a differentiable function that satisfies, for some $0 < \varepsilon' < 1$,

$$\text{i)} \lim_{n\to\infty} \sup_{0\le y\le(1-\varepsilon')R} \left|\frac{\hat{\mu}_n(y)}{\mu(\mathcal{B}_\infty\left(y\right))} - 1\right| = 0 \quad \text{and}$$

$$\text{ii)} \lim_{n\to\infty} \sup_{0\le y\le(1-\varepsilon')R} \left|\frac{\hat{\mu}_n'(y)}{\mu(\mathcal{B}_\infty\left(y\right))} - \frac{1}{2}\right| = 0.$$

Then the following holds for $C > 0$ large enough:

1. *For any continuous function $h : \mathbb{R}_+ \to \mathbb{R}$, such that $h(y) = O\left(e^{\beta y}\right)$ as $y \to \infty$ for some $\beta < \alpha$,*
$$\int_{\mathbb{R}_+\setminus(a_n^-,a_n^+)} h(y)\hat{\rho}_n(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y = O\left(k_n^{-(1+C^2)/2}\right), \tag{4.43}$$
as $n \to \infty$.

2. *If in addition, $h(a_n) \sim h(b_n)$ whenever $a_n \sim b_n$, as $n \to \infty$, then,*
$$\int_0^\infty h(y)\hat{\rho}_n(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y \sim \int_0^\infty h(y)\rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y, \tag{4.44}$$
as $n \to \infty$.

Proof. For simplicity we write $\mu(y) := \mu\left(\mathcal{B}_\infty\left(y\right)\right) = \xi e^{y/2}$ throughout the proof. Observe that $\mu'(y) = \mu(y)/2$ and for the inverse function, $\mu^{-1}(yz) = \mu^{-1}(y) + \mu^{-1}(z)$.

Proof of statement 1. Let $0 < \varepsilon < 1$ and let $\hat{\mu}_n(y)$ be such that conditions i) and ii) hold. Take any $0 < \delta < \min\{\varepsilon, \varepsilon'/3, 1/3\} < 1/2$ and let $Q > 0$ be such that $\hat{\mu}_n(Q) \geq \xi$. We first show that we can restrict to integration over $(Q, (1-\delta)R)$, starting with showing that the integration over $(1-\delta)R \leq y \leq R$ is negligible.

By construction $\delta < \varepsilon'$, and hence by condition i) we have that $\hat{\mu}_n((1-\delta)R) = \Theta\left(\mu((1-\delta)R) = \Theta\left(n^{(1-\delta)}\right)\right.$. Therefore, since $\delta < \varepsilon$ and $k_n = O\left(n^{1-\varepsilon}\right)$, it follows that $\hat{\mu}_n((1-\delta)R)/k_n = \omega\left(n^{\varepsilon-\delta}\right)$ as $n \to \infty$. In particular, $\hat{\mu}((1-\delta)R) = \omega(k_n)$ and hence $\hat{\rho}_n(y, k_n) \leq \hat{\rho}((1-\delta)R, k_n)$ for all $y \geq (1-\delta)R$. It now follows that

$$\begin{aligned}
\int_{(1-\delta)R}^{R} h(y)\hat{\rho}_n(y,k_n)e^{-\alpha y}\,\mathrm{d}y &= O\left(1\right)\hat{\rho}_n((1-\delta)R,k_n)e^{-(\alpha-\beta)(1-\delta)R}\\
&= O\left(\hat{\rho}_n((1-\delta)R,k_n)n^{-2(\alpha-\beta)(1-\delta)}\right),
\end{aligned}$$

where we have used that $h(y) = O\left(e^{\beta y}\right)$. Using Stirling's approximation to bound $\hat{\rho}_n((1-\delta)R, k_n)$ we get

$$\begin{aligned}
\hat{\rho}_n((1-\delta)R,k_n) &= \mathbb{P}\left(\mathrm{Po}(\hat{\mu}_n((1-\delta)R)) = k_n\right)\\
&= \frac{\hat{\mu}_n((1-\delta)R)^{k_n}}{k_n!}e^{-\hat{\mu}_n((1-\delta)R)}\\
&= O\left(1\right)k_n^{-1/2}\left(\frac{\hat{\mu}_n((1-\delta)R)}{k_n}\right)^{k_n}e^{k_n-\hat{\mu}_n((1-\delta)R)}\\
&= O\left(1\right)k_n^{-1/2}e^{k_n\left(1-\frac{\hat{\mu}_n((1-\delta)R)}{k_n}+\log\left(\frac{\hat{\mu}_n((1-\delta)R)}{k_n}\right)\right)}\\
&\leq O\left(1\right)k_n^{-1/2}e^{-\hat{\mu}_n((1-\delta)R)/2},
\end{aligned}$$

where the last line follows since $1 - x + \log(x) \leq -x/2$ for large enough x and $\hat{\mu}_n((1-\delta)R)/k_n \to \infty$. We thus conclude that, for any $C > 0$,

$$\int_{(1-\delta)R}^{R} h(y)\hat{\rho}_n(y,k_n)e^{-\alpha y}\,\mathrm{d}y = O\left(k_n^{-1/2}n^{-2(\beta-\alpha)(1-\delta)}e^{-n^{(1-\delta)}}\right) = O\left(k_n^{-(1+C^2)/2}\right).$$

For the range $(0, Q)$ we have, for any $C > 0$,

$$\begin{aligned}
\int_{0}^{Q} h(y)\hat{\rho}_n(y,k_n)e^{-\alpha y}\,\mathrm{d}y &= O\left(1\right)\hat{\rho}_n(Q,k_n)\\
&= O\left(1\right)k_n^{-1/2}e^{-k_n(\log(k_n)-\hat{\mu}_n(Q))} = O\left(k_n^{-(1+C^2)/2}\right).
\end{aligned}$$

We are thus left to show that for sufficiently large $C > 0$,

$$\int_{\delta}^{a_n^-} \hat{\rho}_n(y,k_n)e^{(\beta-\alpha)y}\,\mathrm{d}y = O\left(k_n^{-(1+C^2)/2}\right), \tag{4.45}$$

and

$$\int_{a_n^+}^{(1-\varepsilon')R} \hat{\rho}_n(y,k_n)e^{(\beta-\alpha)y}\,\mathrm{d}y = O\left(k_n^{-(1+C^2)/2}\right). \tag{4.46}$$

To prove this we first establish a result that will also help with proving statement 2.

Let $(a, b) \subseteq (Q, (1-\delta)R)$ and consider the change of variable $z = \mu^{-1}(\hat{\mu}_n(y))$. Then, writing $\hat{a} = \hat{\mu}_n^{-1}(\mu(a))$ and similar for $\hat{b}$, we get

$$\begin{aligned}\int_a^b \rho(z, k_n) e^{-\alpha z} \, \mathrm{d}z &= \int_a^b \mathbb{P}\left(\mathrm{Po}(\mu(z)) = k_n\right) e^{-\alpha z} \, \mathrm{d}z \\ &= \int_{\hat{a}}^{\hat{b}} \mathbb{P}\left(\mathrm{Po}(\hat{\mu}_n(y)) = k_n\right) e^{-\alpha \mu^{-1}(\hat{\mu}_n(y))} \frac{\hat{\mu}_n'(y)}{\mu'(\mu^{-1}(\hat{\mu}_n(y)))} \, \mathrm{d}y,\end{aligned}$$

where the fraction in the last line follows from the chain rule and the fact that $(\mu^{-1})'(t) = (\mu'(\mu^{-1}(t)))^{-1}$.

Now recall that $\hat{\mu}_n(y)$ satisfies conditions i) and ii). Since $\mu'(y) = \mu(y)/2$ it follows that, uniformly on (a, b),

$$\frac{\hat{\mu}_n'(y)}{\mu'(\mu^{-1}(\hat{\mu}_n(y)))} = \frac{2\hat{\mu}_n'(y)}{\hat{\mu}_n(y)} = \frac{(1 + o(1))2\hat{\mu}_n'(y)}{(1 + o(1))\mu(y)} = (1 + o(1)).$$

Moreover, using $\mu^{-1}(yz) = \mu^{-1}(y) + \mu^{-1}(z)$ and $\mu^{-1}(1) = 0$, we have

$$e^{-\alpha \mu^{-1}(\hat{\mu}_n(y))} = e^{-\alpha(y + \mu^{-1}(1 + o(1)))} = (1 + o(1)) e^{-\alpha y},$$

uniformly on (a, b). These results then imply

$$\int_a^b \rho(z, k_n) e^{-\alpha z} \, \mathrm{d}z = (1 + o(1)) \int_{\hat{a}}^{\hat{b}} \hat{\rho}_n(y, k_n) e^{-\alpha y} \, \mathrm{d}y. \tag{4.47}$$

Now, using (4.47) with $a = \mu^{-1}(\hat{\mu}_n(Q))$ and $b = \mu^{-1}(\hat{\mu}_n(a_n^-)$ we get

$$\int_a^b \rho(y, k_n) e^{(\beta - \alpha) y} \, \mathrm{d}y = (1 + o(1)) \int_Q^{a_n^-} \hat{\rho}_n(y, k_n) e^{(\beta - \alpha) y} \, \mathrm{d}y.$$

Since $\mu^{-1}(\hat{\mu}_n(a_n^-)) = (1 + o(1)) a_n^-$ and $\mu^{-1}(\hat{\mu}_n(a_n^-)) \geq 0$, the left hand side is

$$O(1) \int_0^{a_n^-} \rho(y, k_n) e^{(\beta - \alpha) y} \, \mathrm{d}y,$$

and hence (4.45) follows from Lemma 4.4.1. The proof of (4.46) follows in a similar way.

Proof of statement 2. The proof of the second statement follows the same line of reasoning as above, using (4.47). First, the mean value theorem for definite integrals implies that

$$\int_{a_n^-}^{a_n^+} \hat{\rho}_n(y, k_n) h(y) e^{-\alpha y} \, \mathrm{d}y = h(c_n) \int_{a_n^-}^{a_n^+} \hat{\rho}_n(y, k_n) e^{-\alpha y} \, \mathrm{d}y,$$

for some $c_n \in (a_n^-, a_n^+)$. Since $c_n \sim 2\log(k_n/\xi)$, $h(c_n) \sim h(2\log(k_n/\xi))$ by assumption on h. Therefore it is enough to show that

$$\int_{a_n^-}^{a_n^+} \hat{\rho}_n(y,k_n)e^{-\alpha y}\,\mathrm{d}y = (1+o(1))\int_{a_n^-}^{a_n^+} \rho(z,k_n)e^{-\alpha z}\,\mathrm{d}z.$$

This however follows immediately from (4.47) by picking $a = \mu^{-1}(\hat{\mu}_n(a_n^-))$ and $b = \mu^{-1}(\hat{\mu}_n(a_n^+))$. □

To apply Lemma 4.4.7 to $\rho_{\mathrm{Po}}(y,k_n)$ or $\rho_{\mathrm{box}}(y,k_n)$ we need to show that both these functions satisfy the conditions i) and ii) in Lemma 4.4.7. We will do this in the next two sections and establish that the results from Corollaries 4.4.2, 4.4.3 and 4.4.4 hold when we replace $\rho(y,k_n)$ by either $\rho_{\mathrm{box}}(y,k_n)$ or $\rho_{\mathrm{Po}}(y,k_n)$.

4.4.3 Concentration of heights for the finite box model

The following lemma immediately implies that $\mu\left(\mathcal{B}_{\mathrm{box}}\left(y\right)\right)$ satisfies the conditions of Lemma 4.4.7:

Lemma 4.4.8. *For all $y > 2\log(\pi/2)$,*

$$\mu\left(\mathcal{B}_{box}\left(p\right)\right) = \mu\left(\mathcal{B}_{\infty}\left(p\right)\right)\left(1-\phi_n(y)\right),$$

where $\phi_n(y) \geq 0$ is given by

$$\begin{aligned}\phi_n(y) = &\left(\frac{\pi}{2}\right)^{-(2\alpha-1)} e^{-(\alpha-\frac{1}{2})(R-y)}\\ &- \frac{(2\alpha-1)\pi}{4\alpha}\left(\left(\frac{\pi}{2}\right)^{-2\alpha} e^{-(\alpha-\frac{1}{2})(R-y)} - e^{-(\alpha-\frac{1}{2})R-\frac{y}{2}}\right).\end{aligned}$$

On the other hand, if $y \leq 2\log(\pi/2)$ then

$$\mu(\mathcal{B}_{box}\left(p\right)) = \mu(\mathcal{B}_{\infty}\left(p\right))\left(1-e^{-(\alpha-\frac{1}{2})R}\right).$$

Proof. First note that since we have identified the boundaries of $[-\frac{\pi}{2}e^{\frac{R}{2}}, \frac{\pi}{2}e^{\frac{R}{2}}]$, we can assume, without loss of generality, that $p = (0,y)$. We then have that the boundaries of $\mathcal{B}_{\mathrm{box}}\left(p\right)$ are given by the equations $x' = \pm e^{\frac{y+y'}{2}}$, which intersect the left and right boundaries of $[-\frac{\pi}{2}e^{\frac{R}{2}}, \frac{\pi}{2}e^{\frac{R}{2}}]$ at height

$$h(y) = R + 2\log\left(\frac{\pi}{2}\right) - y.$$

Therefore, if $y \leq 2\log(\pi/2)$ this intersection occurs above the height R of the box $\mathcal{R}$ while in the other case the full region of the box above $h(y)$ is connected to p.

We will first consider the case where $y > 2\log(\pi/2)$. Recall that $\mu(\mathcal{B}_\infty(p)) = \xi e^{\frac{y}{2}}$ where $\xi = \frac{4\alpha\nu}{(2\alpha-1)\pi}$. Also recall that $f_{\alpha,\nu}$ denotes the intensity function of the infinite limit model, as defined in (1.3). Then, after some simple algebra,

$$\begin{aligned}
\mu(\mathcal{B}_{\text{box}}(p)) &= \int_0^{h(y)} \int_{-\frac{\pi}{2}e^{\frac{R}{2}}}^{\frac{\pi}{2}e^{\frac{R}{2}}} \mathbb{1}_{\left\{|x'|\le e^{\frac{y+y'}{2}}\right\}} f_{\alpha,\nu}(x',y')\,dx'\,dy' \\
&\quad + \int_{h(y)}^{R} \int_{-\frac{\pi}{2}e^{\frac{R}{2}}}^{\frac{\pi}{2}e^{\frac{R}{2}}} f_{\alpha,\nu}(x',y')\,dx'\,dy' \\
&= \frac{2\alpha\nu}{\pi} e^{\frac{y}{2}} \int_0^{h(y)} e^{-(\alpha-\frac{1}{2})y'}\,dy' + \alpha\nu e^{\frac{R}{2}} \int_{h(y)}^{R} e^{-\alpha y'}\,dy' \\
&= \xi e^{\frac{y}{2}} \left(1 - \left(\frac{\pi}{2}\right)^{-(2\alpha-1)} e^{-(\alpha-\frac{1}{2})(R-y)}\right) \\
&\quad + \nu e^{\frac{R}{2}} \left(\left(\frac{\pi}{2}\right)^{-2\alpha} e^{-\alpha(R-y)} - e^{-\alpha R}\right) \\
&= \mu(\mathcal{B}_\infty(p))\,(1 - \phi_n(y)) .
\end{aligned}$$

Since, for all $\alpha > \frac{1}{2}$, we have $\frac{2\alpha-1}{2\alpha} \le 1$, and hence,

$$\left(\frac{\pi}{2}\right)^{-(2\alpha-1)} = \left(\frac{\pi}{2}\right)^{-2\alpha} \frac{\pi}{2} \ge \frac{(2\alpha-1)\pi}{4\alpha} \left(\frac{\pi}{2}\right)^{-2\alpha},$$

it follows that $\phi_n(y) \ge 0$.

When $y \le 2\log(\pi/2)$ we have

$$\begin{aligned}
\mu(\mathcal{B}_{\text{box}}(p)) &= \int_0^{R} \int_{-\frac{\pi}{2}e^{\frac{R}{2}}}^{\frac{\pi}{2}e^{\frac{R}{2}}} \mathbb{1}_{\left\{|x'|\le e^{\frac{y+y'}{2}}\right\}} f_{\alpha,\nu}(x',y')\,dx'\,dy' \\
&= \frac{2\alpha\nu}{\pi} e^{\frac{y}{2}} \int_0^{R} e^{-(\alpha-\frac{1}{2})y'}\,dy' \\
&= \mu(\mathcal{B}_\infty(p)) \left(1 - e^{-(\alpha-\frac{1}{2})R}\right) .
\end{aligned}$$

□

From the definition of $\phi_n(y)$ in Lemma 4.4.8 it is immediate that $\rho_{\text{box}}(y,k)$ satisfies the conditions for $\hat{\rho}_n(y,k)$ in Lemma 4.4.7. We thus have the following corollary:

Corollary 4.4.9. *The statements in Corollaries 4.4.2, 4.4.3 and 4.4.4 hold when we replace $\rho(y,k_n)$ with $\rho_{box}(y,k_n)$.*

4.4.4 Concentration of heights for the KPKVB model

We will now show that a concentration of heights argument also applies to the KPKVB model. Due to the hyperbolic distance formula, the computations are however more involved than for the finite box model. Recall that under the coupling between the hyperbolic random graph and the finite box model, for two points p, p' with $y + y' < R$, $p' \in \mathcal{B}(p)$ exactly when $|x - x'|_{\pi e^{R/2}} \leq \Phi(y, y')$ where $|x|_m = \min(|x|, m - |x|)$ is the modular absolute value and $\Phi(y, y')$ is as defined in (1.8). In this setting, the coupling lemma (Lemma 1.6.2) gives that

$$e^{\frac{1}{2}(y+y')} - Ke^{\frac{3}{2}(y+y')-R} \leq \Phi(y, y') \leq e^{\frac{1}{2}(y+y')} + Ke^{\frac{3}{2}(y+y')-R},$$

for some constant K. This result enables us to determine the measure of a ball around a given point $p = (0, y)$. Recall that the hyperbolic ball $\mathcal{B}(p)$ is a subset of $\mathcal{R}$ and not of the hyperbolic disk $\mathcal{D}_R$, i.e. the balls $\mathcal{B}(p)$ "live" in the finite box and not on the hyperbolic disk.

A direct consequence of Lemma 3.3.4 is that $\mu(\mathcal{B}(y)) = \mu(\mathcal{B}_\infty(y))(1 + \phi_n(y))$, where $\phi_n(y) := \mu(\mathcal{B}(y))/\mu(\mathcal{B}_\infty(y)) - 1$ satisfies condition i) in Lemma 4.4.7. To show that condition ii) is also satisfied we need to analyze

$$\phi_n'(y) = \mu(\mathcal{B}_\infty(y))^{-1} \frac{\partial}{\partial y}\mu(\mathcal{B}(y)) - \frac{1}{2}\frac{\mu(\mathcal{B}(y))}{\mu(\mathcal{B}_\infty(y))}, \tag{4.48}$$

where we have used that $\frac{\partial}{\partial y}\mu(\mathcal{B}_\infty(y)) = \frac{1}{2}\mu(\mathcal{B}_\infty(y))$. Again, Lemma 3.3.4 implies that

$$\lim_{n\to\infty} \sup_{0 \leq y \leq (1-\varepsilon)R} \left|\frac{1}{2}\frac{\mu(\mathcal{B}(y))}{\mu(\mathcal{B}_\infty(y))} - \frac{1}{2}\right| = 0.$$

Then, Lemma 3.3.5 shows that the same holds for the first term in (4.48), from which we conclude that $\phi_n(y)$ satisfies condition ii) in Lemma 4.4.7.

We now conclude that similarly to Corollary 4.4.9, the concentration of heights results also hold for $\rho_{\mathrm{Po}}(y, k_n)$:

Corollary 4.4.10. *The statements in Corollaries 4.4.2, 4.4.3 and 4.4.4 hold when we replace $\rho(y, k_n)$ by $\rho_{\mathrm{Po}}(y, k_n)$.*

Remark 4.4.11 (Generalized concentration of heights arguments)**.** Since the results from Corollaries 4.4.2, 4.4.3 and 4.4.4 hold for any of the three functions $\rho(y, k_n)$, $\rho_{\mathrm{box}}(y, k_n)$ and $\rho_{\mathrm{Po}}(y, k_n)$, we will refer only to one of these three when using a concentration of heights argument for any of the three models G_∞, G_{box} and G_{Po}.

4.5 From G_{box} to G_∞ (Proving Proposition 4.3.6)

In this section we shall relate the clustering in the finite box model G_{box} to that of the infinite model. The main goal is to prove Proposition 4.3.6 which states that

$$|\mathbb{E}[c^*(k_n; G_{\mathrm{box}})] - \gamma(k_n)| = o(s(k_n)).$$

Recall that G_{box} is obtained by restricting the Poisson point process $\mathcal{P}_{\alpha,\nu}$ to the box $\mathcal{R} = (-I_n, I_n] \times (0, R]$, with $I_n = \frac{\pi}{2}e^{R/2}$ and connecting two points $p_1, p_2 \in \mathcal{R}$ if and only if $|x_1 - x_2|_{\pi e^{R/2}} \leq e^{(y_1+y_2)/2}$. We also recall that by definition of the norm $|.|_{\pi e^{R/2}}$ the left and right boundaries of $\mathcal{R}$ are identified. See Section 1.6.2 for more details. Due to this identification of the boundaries some triples of vertices that form a triangle in the finite box model do not form a triangle in the infinite model. Therefore, to establish the required result we need to compute the asymptotic difference between triangle counts in both models. To keep notation concise we write $|\cdot|_n$ for the norm $|\cdot|_{\pi e^{R/2}}$.

For any $p \in \mathbb{R} \times \mathbb{R}_+$ we define for the finite box model,

$$T_{\text{box}}(p) = \sum_{p_1, p_2 \in \mathcal{P} \setminus \{p\}}^{\neq} T_{\text{box}}(p, p_1, p_2),$$

where the sum is over all distinct pairs in $\mathcal{P} \setminus p$ and

$$T_{\text{box}}(p, p_1, p_2) = \mathbb{1}_{\{p_1 \in \mathcal{B}_{\text{box}}(p)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\text{box}}(p)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\text{box}}(p_1)\}}.$$

Similarly, for the infinite model we define

$$T_\infty(y) = \sum_{p_1, p_2 \in \mathcal{P} \setminus \{(0,y)\}}^{\neq} T_\infty(y, p_1, p_2),$$

where

$$T_\infty(y, p_1, p_2) = \mathbb{1}_{\{p_1 \in \mathcal{B}_\infty(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_\infty(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_\infty(p_1)\}}.$$

Here we slightly abuse notation and write $\mathcal{B}_\infty(y)$ for $\mathcal{B}_\infty((0, y))$. We will adopt this notational convention throughout the remainder of this section, to keep notation concise. We further write $N_{\text{box}}(k)$ to denote the number of vertices with degree k in G_{box}.

We will first relate $\gamma(k_n)$ and $\mathbb{E}[c^*(k_n; G_{\text{box}})]$ using $T_\infty(y)$ and $T_{\text{box}}(y)$. Recall the definition of $\mathcal{K}_C(k_n)$ as

$$\mathcal{K}_C(k_n) = \left\{ y \in \mathbb{R}_+ : \frac{k_n - C\sqrt{k_n \log(k_n)}}{\xi} \leq e^{\frac{y}{2}} \leq \frac{k_n + C\sqrt{k_n \log(k_n)}}{\xi} \right\}.$$

Lemma 4.5.1. *Let $\gamma(k_n)$ be defined as in (4.6). Then as $n \to \infty$,*

$$\gamma(k_n) = (1 + o(1)) \frac{1}{k_n^2 p_{k_n}} \int_{\mathcal{K}_C(k_n)} \mathbb{E}[T_\infty(y)] \rho(y, k) \alpha e^{-\alpha y} \, dy. \tag{4.49}$$

Moreover,

$$\mathbb{E}[c^*(k_n; G_{\text{box}})] = (1 + o(1)) \frac{1}{k_n^2 p_{k_n}} \int_{\mathcal{K}_C(k_n)} \mathbb{E}[T_{box}(y)] \rho(y, k_n) \alpha e^{-\alpha y} \, dy \tag{4.50}$$

as $n \to \infty$.

Proof. Recall that

$$P(y) = \mathbb{E}\left[\mathbb{1}_{\{u_1 \in \mathcal{B}_\infty(u_2)\}}\right],$$

where u_1 and u_2 are independent and distributed according to the probability density $\mu\left(\mathcal{B}_\infty\left(y\right)\right)^{-1} \mathbb{1}_{\{u_i \in \mathcal{B}_\infty(y)\}} f(x_i, y_i)$. It then follows from the Campbell-Mecke formula that

$$\begin{aligned}
\mathbb{E}\left[T_\infty(y)\right] &= \int \mathbb{1}_{\{p_1 \in \mathcal{B}_\infty(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_\infty(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_\infty(p_1)\}} f(x_1, y_1) f(x_2, y_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \, \mathrm{d}y_1 \, \mathrm{d}y_2 \\
&= \mu\left(\mathcal{B}_\infty\left(y\right)\right)^2 P(y).
\end{aligned}$$

It then follows that,

$$\begin{aligned}
\gamma(k_n) &= \frac{1}{p_{k_n}} \cdot \int_0^\infty P(y) \rho(y, k) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&= \frac{1}{p_{k_n}} \int_0^\infty \mathbb{E}\left[T_\infty(y)\right] \mu\left(\mathcal{B}_\infty\left(y\right)\right)^{-2} \rho(y, k) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&= (1 + o\left(1\right)) \frac{1}{k_n^2 p_{k_n}} \int_0^\infty \mathbb{E}\left[T_\infty(y)\right] \rho(y, k) \alpha e^{-\alpha y} \, \mathrm{d}y,
\end{aligned}$$

where the last line is due to a concentration of heights argument, which yields that $\mu\left(\mathcal{B}_\infty\left(y\right)\right) = (1 + o(1)) k_n$.

For (4.50) we recall that

$$c^*(k_n; G_{\text{box}}) = \frac{1}{\mathbb{E}\left[N_{\text{box}}(k_n)\right]} \sum_{p \in \mathcal{P}} c_{\text{box}}(p) \mathbb{1}_{\{\deg_{\text{box}}(p) = k_n\}},$$

where $c_{\text{box}}(p)$ can be expressed as

$$c_{\text{box}}(p) = \frac{1}{\binom{\deg_{\text{box}}(p)}{2}} \sum_{p_1, p_2 \in \mathcal{P} \setminus \{p\}}^{\neq} T_{\text{box}}(p, p_1, p_2) = \frac{T_{\text{box}}(p)}{\binom{\deg_{\text{box}}(p)}{2}}.$$

Recall that f denotes the intensity function as defined in (1.3). By the Campbell-Mecke formula

$$\begin{aligned}
\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right] &= \frac{1}{\mathbb{E}\left[N_{\text{box}}(k_n)\right]} \int_{\mathcal{R}} \mathbb{E}\left[c_{\text{box}}(p) \mathbb{1}_{\{\deg_{\text{box}}(p) = k_n\}}\right] f(x, y) \, \mathrm{d}x \, \mathrm{d}y \\
&= \frac{1}{\mathbb{E}\left[N_{\text{box}}(k_n)\right]} \int_{\mathcal{R}} \mathbb{E}\left[c_{\text{box}}(p) | \deg_{\text{box}}(p) = k_n\right] \rho_{\text{box}}(p, k_n) f(x, y) \, \mathrm{d}x \, \mathrm{d}y \\
&= (1 + o\left(1\right)) \frac{n}{\mathbb{E}\left[N_{\text{box}}(k_n)\right]} \\
&\qquad \times \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[c_{\text{box}}(y) | \deg_{\text{box}}(y) = k_n\right] \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y,
\end{aligned}$$

where the last line follows from a concentration of heights argument, for which we have used that $\mathbb{E}\left[c_{\text{box}}(y) | \deg_{\text{box}}(y) = k_n\right] \leq 1$. To analyze the conditional expectation we observe that, similarly to the analysis of $\gamma(k_n)$, conditioned on there being k_n points in $\mathcal{B}_{\text{box}}(y)$, each point $u_i = (x_i, y_i)$ is independently distributed according to the probability density $\mu(\mathcal{B}_{\text{box}}(y))^{-1} \mathbb{1}_{\{u_i \in \mathcal{B}_{\text{box}}(y)\}} f(x_i, y_i)$. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[c_{\text{box}}(y) | \deg_{\text{box}}(y) = k_n\right] \\
&= \binom{k_n}{2}^{-1} \mathbb{E}\left[\sum_{1 \le i < j \le k_n} \mathbb{1}_{\{u_i \in \mathcal{B}_{\text{box}}(u_j)\}}\right] \\
&= \mathbb{P}\left(u_1 \in \mathcal{B}_{\text{box}}(u_2)\right) \\
&= \mu(\mathcal{B}_{\text{box}}(y))^{-2} \iint T_{\text{box}}(y, p_1, p_2) f(x_1, y_1) f(x_2, y_2) \, \mathrm{d}x_1 \, \mathrm{d}y_1 \, \mathrm{d}x_2 \, \mathrm{d}y_2 \\
&= \mu(\mathcal{B}_{\text{box}}(y))^{-2} \, \mathbb{E}\left[T_{\text{box}}(y)\right].
\end{aligned}$$

and thus, by applying a concentration of heights argument on $\mu(\mathcal{B}_{\text{box}}(y))^{-2}$,

$$\begin{aligned}
\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right] = (1 + o(1)) &\frac{n\mu(\mathcal{B}_{\text{box}}(2\log(k_n/\xi)))^{-2}}{\mathbb{E}\left[N_{\text{box}}(k_n)\right]} \\
&\times \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[T_{\text{box}}(y)\right] \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y.
\end{aligned}$$

To finish the argument, we first note that $\mu(\mathcal{B}_{\text{box}}(2\log(k_n/\xi)))^{-2} = (1 + o(1)) k_n^2$, while

$$\mathbb{E}\left[N_{\text{box}}(k_n)\right] = \int_{\mathcal{R}} \rho_{\text{box}}(y, k_n) f(x, y) \, \mathrm{d}x \, \mathrm{d}y,$$

so that by a concentration of heights argument,

$$\mathbb{E}\left[N_{\text{box}}(k_n)\right] = (1 + o(1)) n \int_0^\infty \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y = (1 + o(1)) n p_{k_n}.$$

We therefore conclude that

$$\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right] = (1 + o(1)) \frac{1}{k_n^2 p_{k_n}} \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[T_{\text{box}}(y)\right] \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y.$$

$\square$

Comparing (4.49) and (4.50), we conclude that to prove Proposition 4.3.6 it is enough to show that

$$\left| \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[T_{\text{box}}(y) - T_\infty(y)\right] \rho(y, k) \alpha e^{-\alpha y} \, dy \right| = o\left(s(k_n) p_{k_n} k_n^2\right), \tag{4.51}$$

which means we have to compute the expected difference in triangles between both models.

4.5.1 Comparing triangles between G_∞ and G_{box}

To analyze $T_{\text{box}}(y_0) - T_\infty(y_0)$ we first reiterate that the difference between the indicator $\mathbb{1}_{\{p_1 \in \mathcal{B}_{\text{box}}(p)\}}$ in the finite box model and $\mathbb{1}_{\{p_1 \in \mathcal{B}_\infty(p)\}}$ is that in G_{box} we have identified the boundaries of the interval $[-\frac{\pi}{2}e^{R/2}, \frac{\pi}{2}e^{R/2}]$ and we stop at height $y = R$. This induces a difference in triangle counts between both models. To see this, note that for any $p = (x, y)$ with $0 \le y \le R$ we have that $\mathcal{B}_{\text{box}}(p) = \mathcal{B}_\infty(p) \cap \mathcal{R}$. This means that if $p', p_2 \in \mathcal{B}_{\text{box}}(p)$ and $p_2 \in \mathcal{B}_\infty(p') \cap \mathcal{R}$ then $p_2 \in \mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')$ and hence (p, p', p_2) form a triangle both in G_{box} and G_∞. However, it could happen that there are points in the intersection $\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')$ that are not in $\mathcal{B}_\infty(p) \cap \mathcal{B}_\infty(p')$. Let us denote this region by $\mathcal{T}(p, p')$, see Figure 4.3 for an example of this region. Then, any $p_2 \in \mathcal{T}(p, p')$ creates a triangle with p and p' in G_{box} that is not present in G_∞. Finally, any point $p_2 \in \mathcal{B}_\infty(p) \cap \mathcal{B}_\infty(p')$ with height $y_2 > R$ creates a triangle with p, p' in G_∞ but not in G_{box}.

Let us now define the triangle count function

$$\widetilde{T}_{\text{box}}(p_0) = \sum_{(p_1,p_2) \in \mathcal{P} \setminus \{p_0\}}^{\neq} \widetilde{T}_{\text{box}}(p_0, p_1, p_2),$$

where

$$\widetilde{T}_{\text{box}}(p_0, p_1, p_2) = \mathbb{1}_{\{p_1 \in \mathcal{B}_{\text{box}}(p)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\text{box}}(p)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_\infty(p_1) \cap \mathcal{R}\}}.$$

Then $\widetilde{T}_{\text{box}}(p_0)$ only counts those triangles attached to p_0 that exist in both G_{box} and G_∞ and thus, by definition of the region $\mathcal{T}(p_0, p_1)$,

$$T_{\text{box}}(p_0) - \widetilde{T}_{\text{box}}(p_0) = \sum_{p_1, p_2 \in \mathcal{P} \setminus \{p_0\}}^{\neq} \mathbb{1}_{\{p_1 \in \mathcal{B}_{\text{box}}(p_0)\}} \mathbb{1}_{\{p_2 \in \mathcal{T}(p_0, p_1)\}}.$$

The next result, which is crucial for the proof of Proposition 4.3.6, computes the expected measure of $\mathcal{T}(p, p')$ with respect to p':

Lemma 4.5.2. *Let $p_0 = (0, y)$ with $y \in \mathcal{K}_C(k_n)$. Then as $n \to \infty$,*

$$\mathbb{E}\left[\left|T_{box}(p_0) - \widetilde{T}_{box}(p_0)\right|\right] = y\, O\left(n^{-(2\alpha-1)}\right) + e^y\, O\left(n^{-(4\alpha-2)}\right).$$

The proof of the lemma is not difficult but cumbersome, since it involves computing many different integrals. We postpone this proof till the end of this section and proceed with the main goal, proving Proposition 4.3.6.

First we state a small lemma about the scaling of $s(k_n)$ that will be very useful:

Lemma 4.5.3. *Let $s(k_n)$ be as defined in (4.33). Then for any $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$, as $n \to \infty$,*

$$n^{-(2\alpha-1)} = o\left(s(k_n)\right).$$

Proof. First let $\frac{1}{2} < \alpha < \frac{3}{4}$. Then

$$n^{-(2\alpha-1)} s(k_n)^{-1} = n^{-(2\alpha-1)} k_n^{4\alpha-2} = o\left(n^{-(2\alpha-1)+\frac{4\alpha-2}{2\alpha+1}}\right) = o\left(n^{-\frac{4\alpha^2-4\alpha+1}{2\alpha+1}}\right) = o(1),$$

since $4\alpha^2 - 4\alpha + 1 > 0$ for all $\alpha > \frac{1}{2}$. Similarly, for $\alpha \geq \frac{3}{4}$ we have that $4\alpha^2 > 2$ and hence,

$$n^{-(2\alpha-1)} s(k_n)^{-1} = o\left(n^{-(2\alpha-1)} k_n\right) = o\left(n^{-\frac{4\alpha^2-2}{2\alpha+1}}\right) = o(1).$$

□

We now proceed with the proof of the main result of this section:

Proof of Proposition 4.3.6. Let us write $\mathcal{R}' := (\mathbb{R} \times \mathbb{R}_+) \setminus \mathcal{R}$ and let $p_0 = (0, y)$ denote the typical point. Next we recall that it is enough to show (4.51), so that in particular we have that $y \in \mathcal{K}_C(k_n)$.

Now

$$|T_{\text{box}}(p_0) - T_\infty(p_0)| = \left|T_{\text{box}}(p_0) - \widetilde{T}_{\text{box}}(p_0)\right| + \sum_{p_1, p_2 \in \mathcal{P} \cap \mathcal{R}'}^{\neq} T_\infty(p_0, p_1, p_2),$$

so that by the Campbell-Mecke formula

$$\begin{aligned}|\mathbb{E}[T_{\text{box}}(p_0) - T_\infty(p_0)]| &\leq \mathbb{E}\left[\left|T_{\text{box}}(p_0) - \widetilde{T}_{\text{box}}(p_0)\right|\right] \\ &\quad + \int_{\mathcal{R}'} \int_{\mathcal{R}'} T_\infty(p_0, p_1, p_2) f(x_1, y_1) f(x_2, y_2) \, dx_2 \, dy_2 \, dx_1 \, dy_1.\end{aligned}$$

The first part is taken care of by Lemma 4.5.2. For the other integral we have

$$\begin{aligned}&\iint_{\mathcal{R}'} T_\infty(p_0, p_1, p_2) f(x_1, y_1) f(x_2, y_2) \, dx_2 \, dy_2 \, dx_1 \, dy_1 \\ &\leq \left(\int_{\mathcal{R}'} \mathbb{1}_{\{p_1 \in \mathcal{B}_\infty(p_0)\}} f(x_1, y_1) \, dx_1 \, dy_1\right)^2 = O\left(\left(e^{y/2} \int_R^\infty e^{-(\alpha - \frac{1}{2}) y_1} \, dy_1\right)^2\right) \\ &= O\left(e^y e^{-(2\alpha-1)R}\right) = O\left(e^y n^{-(4\alpha-2)}\right).\end{aligned}$$

Thus we conclude, using Lemma 4.5.2, that

$$|\mathbb{E}[T_{\text{box}}(p_0) - T_\infty(p_0)]| = O\left(y n^{-(2\alpha-1)} + n^{-(4\alpha-2)} e^y\right). \tag{4.52}$$

Therefore, on $\mathcal{K}_C(k_n)$,

$$\int_{\mathcal{K}_C(k_n)} \rho(y_0, k_n) \, |\mathbb{E}[T_{\text{box}}(p_0) - T_\infty(p_0)]| \, e^{-\alpha y_0} \, dy_0$$

$$= O(1)\left(\log(k_n)n^{-(2\alpha-1)} + k_n^2 n^{-(4\alpha-2)}\right)\int_0^\infty \rho(y_0,k_n)e^{-\alpha y_0}\,\mathrm{d}y_0$$
$$= O(1)\left(\log(k_n)n^{-(2\alpha-1)} + k_n^2 n^{-(4\alpha-2)}\right)p_{k_n} = o\left(s(k_n)p_{k_n}k_n^2\right),$$

where the last part follows from Lemma 4.5.3 and the fact that $s(k_n)^2 = o(s(k_n))$. This establishes (4.51) and hence finishes the proof. □

From the proof of Proposition 4.3.6 we obtain the following useful corollary, which will be used in Section 4.6. Recall that

$$\rho_{\text{box}}(y,k_n) = \mathbb{P}\left(\text{Po}(\mu(\mathcal{B}_{\text{box}}(y))) = k_n\right)$$

denotes the degree distribution of a point $p_0 = (0,y)$ in G_{box}.

Corollary 4.5.4. *Let $p_0 = (0,y)$. Then, as $n \to \infty$,*

$$\int_{-I_n}^{I_n}\int_{\mathcal{K}_C(k_n)} \rho_{box}(y,k_n)\mathbb{E}\left[\widetilde{T}_{box}(p_0)\right] f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= (1+o(1))nk_n^2\int_0^\infty P(y)\rho(y,k_n)\alpha e^{-\alpha y}\,\mathrm{d}y.$$

In particular,

$$\int_{\mathcal{K}_C(k_n)} \rho_{box}(y,k_n)\mathbb{E}\left[\widetilde{T}_{box}(p_0)\right] f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \Theta\left(nk_n^{-(2\alpha-1)}s(k_n)\right).$$

Proof. We first write

$$\mathbb{E}\left[\left|\widetilde{T}_{\text{box}}(y) - T_\infty(y)\right|\right] \le \mathbb{E}\left[\left|T_{\text{box}}(y) - \widetilde{T}_{\text{box}}(y)\right|\right] + \mathbb{E}\left[\left|T_{\text{box}}(y) - T_\infty(y)\right|\right].$$

Therefore, Lemma 4.5.2 and (4.52) imply that, uniformly for $y \in \mathcal{K}_C(k_n)$,

$$\mathbb{E}\left[\left|\widetilde{T}_{\text{box}}(y) - T_\infty(y)\right|\right] = O\left(\log(k_n)n^{-(2\alpha-1)} + k_n^2 n^{-(4\alpha-2)}\right) = o\left(s(k_n)k_n^2\right),$$

where the last part is due to Lemma 4.5.3. Next, since $\mathbb{E}\left[T_\infty(y)\right] = \mu\left(\mathcal{B}_\infty(y)\right)^2 P(y)$, we get

$$\mathbb{E}\left[\widetilde{T}_{\text{box}}(y)\right] = \mathbb{E}\left[T_\infty(y)\right] + \mathbb{E}\left[\widetilde{T}_{\text{box}}(y) - T_\infty(y)\right] = k_n^2 P(y) + o\left(s(k_n)k_n^2\right),$$

uniformly on $\mathcal{K}_C(k_n)$. Therefore, we can apply a concentration of heights argument to replace $\rho_{\text{box}}(y,k_n)$ by $\rho(y,k_n)$ and thus obtain

$$\int_{\mathcal{K}_C(k_n)} \rho_{\text{box}}(y,k_n)\mathbb{E}\left[\widetilde{T}_{\text{box}}(y)\right] f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

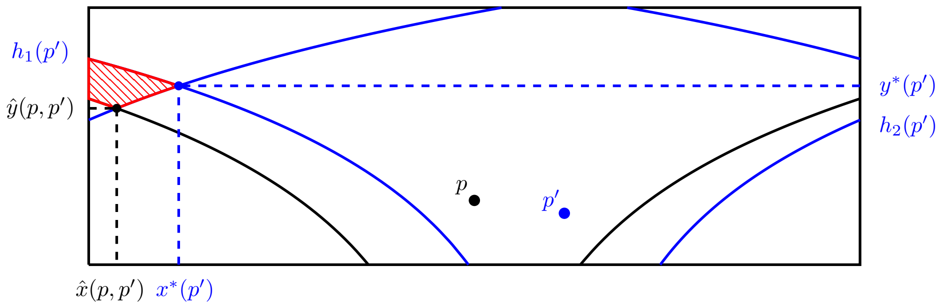

Figure 4.3: Example configuration of two points p and p' for which $\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')$ is not a subset of $\mathcal{B}_{\infty}(p) \cap \mathcal{B}_{\infty}(p')$. The red region indicates the area belonging to $\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')$ but not to $\mathcal{B}_{\infty}(p) \cap \mathcal{B}_{\infty}(p')$.

$$
\begin{aligned}
&= nk_n^2 \int_{\mathcal{K}_C(k_n)} \rho(y, k_n)\left(P(y) + o\left(s(k_n)\right)\right) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&= (1 + o(1)) n k_n^2 \int_{\mathcal{K}_C(k_n)} P(y)\rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&= (1 + o(1)) n k_n^2 \int_0^\infty P(y)\rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y,
\end{aligned}
$$

where we have used that $P(y) = \Theta\left(s(k_n)\right)$ on $\mathcal{K}_C(k_n)$. This proves the first statement. The second statement follows by observing that

$$
\int_0^\infty P(y)\rho(y, k_n)\alpha e^{-\alpha y} \, \mathrm{d}y = p_{k_n}\gamma(k_n) = \Theta\left(k_n^{-(2\alpha+1)} s(k_n)\right).
$$

□

4.5.2 Counting missing triangles

We now come back to computing the expected number of triangles attached to a vertex at height y in G_{box}, which are not present in G_{∞}.

Recall that $\mathcal{T}(p, p')$ denotes the region of points which form triangles with p and p' in G_{box} but not in G_{∞}. Figure 4.3 shows an example of a configuration where $\mathcal{T}(p, p') \neq \emptyset$. We observe that $\mathcal{T}(p, p') \neq \emptyset$ because the right boundary of the ball $\mathcal{B}_{\text{box}}(p')$ exits the right boundary of the box $\mathcal{R}$ and then, since we identified the boundaries, continues from the left so that $\mathcal{B}_{\text{box}}(p')$ covers part of the ball $\mathcal{B}_{\text{box}}(p)$ which would not be covered in the infinite limit model.

To further analyze this, let us introduce some notation. For any $p = (x, y) \in \mathcal{R}$

we will define the left and right boundary functions as, respectively,

$$b_p^-(z) = \begin{cases} 2\log(x-z) - y, & \text{if } -\frac{\pi}{2}e^{R/2} \le z \le x - e^{y/2}, \\ 2\log\left(\pi e^{R/2} + x - z\right) - y, & \text{if } x - e^{(y+R)/2} + \pi e^{R/2} \le z \le \frac{\pi}{2}e^{R/2}, \\ 0, & \text{otherwise,} \end{cases} \tag{4.53}$$

$$b_p^+(z) = \begin{cases} 2\log(z-x) - y, & \text{if } x + e^{y/2} \le z \le \frac{\pi}{2}e^{R/2}, \\ 2\log\left(\pi e^{R/2} + z - x\right) - y, & \text{if } -\frac{\pi}{2}e^{R/2} \le z \le x + e^{(y+R)/2} - \pi e^{R/2}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.54}$$

Note that these functions describe the boundaries of the ball $\mathcal{B}_{\text{box}}(p)$. In particular, $p' = (x', y') \in \mathcal{B}_{\text{box}}(p)$ if and only if $y' \ge \min\left\{b_p^-(x'), b_p^+(x')\right\}$.

Since we have identified the left and right boundary of $\mathcal{R}$ we can assume, without loss of generality, that $x = 0$. Due to symmetry it is then enough to restrict the analysis to the case where $x' > 0$. For this case there are two important points in the box $\mathcal{R}$. These are the intersection between the left boundary of p' and the right boundary of p', as it continues from the left side of the box, and the left boundary of p. We denote by $(x^*(p'), y^*(p'))$ the intersection between the left and right boundary of p' and by $(\hat{x}(p,p'), \hat{y}(p,p'))$ the intersection between the left boundary of p and the right boundary of p', see Figure 4.3.

Let us derive the expressions for the coordinates of these two points, starting with $(x^*(p'), y^*(p'))$. The x-coordinate $x^*(p')$ is the solution to the equation $b_{p'}^+(z) = b_{p'}^-(z)$ for $-\frac{\pi}{2}e^{R/2} \le z \le x + e^{(y+R)/2} - \pi e^{R/2}$. This equation becomes

$$2\log\left(\pi e^{R/2} + z - x'\right) - y' = 2\log(x' - z) - y',$$

whose solution is $x^*(p') := x' - \frac{\pi}{2}e^{R/2}$. Plugging this into either the left or right hand side of the above equation yields the y-coordinate $y^*(p') = 2\log\left(\frac{\pi}{2}e^{R/2}\right) - y'$. In a similar way, the x-coordinate $\hat{x}(p,p')$ is the solution to the equation $b_{p'}^+(z) = b_p^-(z)$ for $-\frac{\pi}{2}e^{R/2} \le z \le x + e^{(y+R)/2} - \pi e^{R/2}$, i.e.

$$2\log\left(\pi e^{R/2} + z - x'\right) - y' = 2\log(x - z) - y.$$

This solution is $\frac{x' - \pi e^{R/2}}{1 + e^{(y'-y)/2}}$ and again $\hat{y}(p,p')$ is obtained by plugging the solution into either the left or right hand side of the equation, yielding $\hat{y}(p,p') = 2\log\left(\frac{\pi e^{R/2} - x'}{e^{y/2} + e^{y'/2}}\right)$.

To summarize we have:

$$x^*(p') = x' - \frac{\pi}{2}e^{R/2},$$
$$y^*(p') = 2\log\left(\frac{\pi}{2}e^{R/2}\right) - y',$$

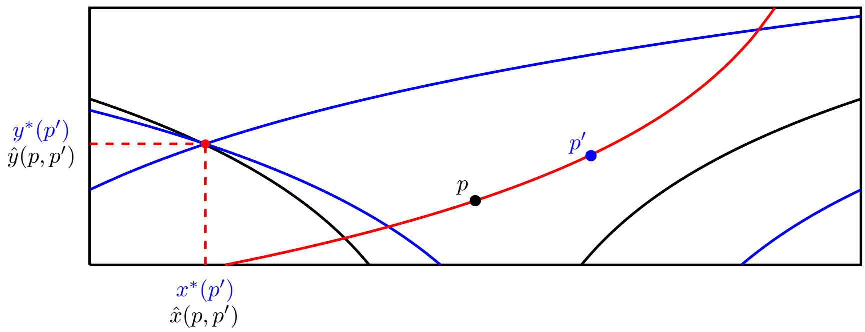

Figure 4.4: Example for a given p of the boundary function $x' \mapsto b_p^*(x')$, given by the red curve, which determines whether $\mathcal{T}(p,p') = \emptyset$ or not. We see that when $y' = b_p^*(x')$ then $(\hat{x}(p,p'), \hat{y}(p,p')) = (x^*(p'), y^*(\prime))$.

$$\hat{x}(p,p') = \frac{x' - \pi e^{R/2}}{1 + e^{(y'-y)/2}},$$
$$\hat{y}(p,p') = 2\log\left(\frac{\pi e^{R/2} - x'}{e^{y/2} + e^{y'/2}}\right).$$

The crucial observation is that $\mathcal{T}(p,p') = \emptyset$ as long as the point $(x^*(p'), y^*(p'))$ is above the left boundary of p. This happens exactly when $y^*(p') > b_p^-(x^*(p'))$. Therefore the boundary of this event is given by the equation $y^*(p') = b_p^-(x^*(p'))$ which reads

$$2\log\left(\frac{\pi}{2}e^{R/2}\right) - y' = 2\log\left(\frac{\pi}{2}e^{R/2} - x'\right) - y.$$

Solving this equation gives the function

$$b_p^*(z) = y - 2\log\left(1 - \frac{z}{\frac{\pi}{2}e^{R/2}}\right), \tag{4.55}$$

which is displayed by the red curve in Figure 4.4. It holds that $y^*(p') > b_p^-(x^*(p'))$ if and only if $y' < b_p^*(x')$ and hence we have that $\mathcal{T}(p,p') = \emptyset$ for all $p' \in \mathcal{R}$ for which $y' \geq b_p^*(x')$. We also note that when $y' = b_p^*(x')$ the two points $(x^*(p'), y^*(p'))$ and $(\hat{x}(p,p'), \hat{y}(p,p'))$ coincide.

This analysis allows us to compute the expected difference in the number of triangles for the finite box model and the infinite model, for a typical vertex with height y, i.e. prove Lemma 4.5.2.

Proof of Lemma 4.5.2. Let $I_n = \frac{\pi}{2}e^{R/2}$. Due to symmetry it is enough to show that

$$\int_0^R \int_0^{I_n} \mu\left(\mathcal{T}(p,p_1)\right) f(x_1,y_1)\, \mathrm{d}x_1\, \mathrm{d}y_1 = O\left(yn^{-(2\alpha-1)} + n^{-(2\alpha-1)}e^y\right). \tag{4.56}$$

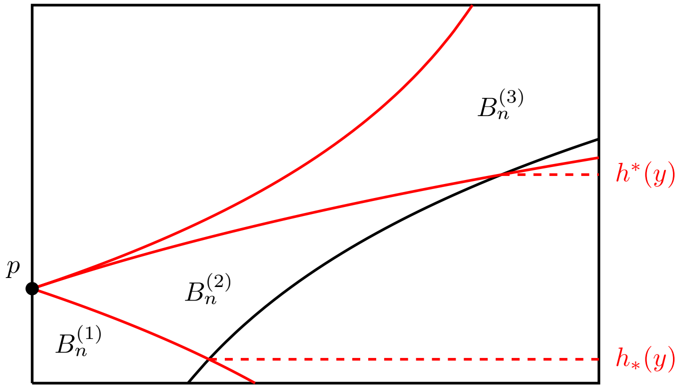

Figure 4.5: Three different areas $B_n^{(i)}$ used in the proof of Lemma 4.5.2.

The proof goes in two stages. First we compute $\mu\left(\mathcal{T}(p,p_1)\right)$ by splitting it over three disjoint regimes with respect to p_1, with $x_1 \geq 0$. Then we do the integration with respect to p_1.

Computing $\boldsymbol{\mu\left(\mathcal{T}(p,p_1)\right)}$

Recall that $I_n = \frac{\pi}{2}e^{R/2}$ and define the sets

$$A_n^{(1)} = \left\{p_1 = (x_1,y_1) \in \mathcal{R} \,:\, 0 \leq y_1 \leq y - 2\log(I_n/(I_n - x_1))\right\},$$
$$A_n^{(2)} = \left\{p_1 = (x_1,y_1) \in \mathcal{R} \,:\, y - 2\log(I_n/(I_n - x_1)) < y_1 \leq y + 2\log\left(1 + \frac{x_1}{I_n}\right)\right\},$$
$$A_n^{(3)} = \left\{p_1 = (x_1,y_1) \in \mathcal{R} \,:\, y + 2\log\left(1 + \frac{x_1}{I_n}\right) < y_1 \leq y + 2\log\left(\frac{I_n}{I_n - x_1}\right)\right\},$$

and let $B_n^{(i)} = \mathcal{B}_{\text{box}}(p) \cap A_n^{(i)}$, for $i = 1,2,3$, see Figure 4.5. Here the heights of the two intersections are given by

$$h_*(y) = y + 2\log\left(\frac{I_n}{I_n + e^y}\right), \tag{4.57}$$
$$h^*(y) = y + 2\log\left(\frac{I_n}{I_n - e^y}\right). \tag{4.58}$$

With these definitions we have that the union $B_n := \bigcup_{i=1}^n B_n^{(i)}$ denotes the area under the red curve in Figure 4.4 and hence, for all $p_1 \in \mathcal{R} \setminus B_n$ with $x_1 \geq 0$ we have that $\mathcal{T}(p,p_1) = \emptyset$. So we only need to consider $p_1 \in B_n$. We shall establish the following result:

$$\mu\left(\mathcal{T}(p,p_1)\right) = \begin{cases} O\left(I_n^{-2\alpha}e^{\alpha y_1}\right) & \text{if } p_1 \in B_n^{(1)}, \\ O\left(I_n^{-2\alpha}e^{\alpha y}\right) & \text{if } p_1 \in B_n^{(2)} \cup B_n^{(3)}. \end{cases} \tag{4.59}$$

Depending on which set p_1 belongs to, the set $\mathcal{T}(p, p_1)$ has a different shape. We displayed these shapes in Figure 4.6 as a visual aid to follow the computations below.

Case $p_1 \in B_n^{(1)}$: $0 \le y_1 \le y - 2\log(I_n/(I_n - x_1))$ In this case the integral over p_2 splits into two parts

$$\mathcal{I}_n^{(1)}(p_1) := \int_{h_2(p_1)}^{y^*(p_1)} \int_{-I_n}^{x_1 + e^{(y_1+y_2)/2} - 2I_n} e^{-\alpha y_2} \, \mathrm{d}x_2 \, \mathrm{d}y_2$$
$$\mathcal{I}_n^{(2)}(p_1) := \int_{y^*(p_1)}^{h_1(p_1)} \int_{x^*(p_1)}^{x_1 - e^{(y_1+y_2)/2}} e^{-\alpha y_2} \, \mathrm{d}x_2 \, \mathrm{d}y_2.$$

We first compute $\mathcal{I}_n^{(1)}$ as

$$\begin{aligned}
\mathcal{I}_n^{(1)}(p_1) &= \int_{h_2(p_1)}^{y^*(p_1)} \left(x_1 + e^{(y_1+y_2)/2} - I_n\right) e^{-\alpha y_2} \, \mathrm{d}y_2 \\
&\le e^{y_1/2} \int_{h_2(p_1)}^{y^*(p_1)} e^{-(\alpha - \frac{1}{2})y_2} \, \mathrm{d}y_2 \\
&= \frac{2e^{y_1/2}}{2\alpha - 1} \left(e^{-(\alpha-\frac{1}{2})h_2(p_1)} - e^{-(\alpha-\frac{1}{2})y^*(p_1)}\right) \\
&= \frac{2e^{\alpha y_1}}{2\alpha - 1} I_n^{-(2\alpha-1)} \left(\left(1 - \frac{x_1}{I_n}\right)^{-(2\alpha-1)} - 1\right) \\
&= O\left(I_n^{-2\alpha} x_1 e^{\alpha y_1}\right),
\end{aligned}$$

where we have used that $x_1 \le e^{(y+y_1)/2} = o(I_n)$ for all $y_1 \le y$ and $y \in \mathcal{K}_C(k_n)$ so that

$$\left(\left(1 - \frac{x_1}{I_n}\right)^{-(2\alpha-1)} - 1\right) = O\left(\frac{x_1}{I_n}\right) \quad \text{as } n \to \infty.$$

For $\mathcal{I}_n^{(2)}(p_1)$ we have

$$\begin{aligned}
\mathcal{I}_n^{(2)}(p_1) &= \int_{y^*(p_1)}^{h_1(p_1)} \left(I_n + x_1 - e^{(y_1+y_2)/2}\right) e^{-\alpha y_2} \, \mathrm{d}y_2 \\
&\le 2I_n \int_{y^*(p_1)}^{h_1(p_1)} e^{-\alpha y_2} \, \mathrm{d}y_2 \\
&= \frac{2}{\alpha} I_n \left(I_n^{-2\alpha} e^{\alpha y_1} - (I_n + x_1)^{-2\alpha} e^{-\alpha y_1}\right) \\
&= O\left(I_n^{-2\alpha} x_1 e^{\alpha y_1}\right) = O\left(I_n^{-(2\alpha-1)} e^{\alpha y_1}\right).
\end{aligned}$$

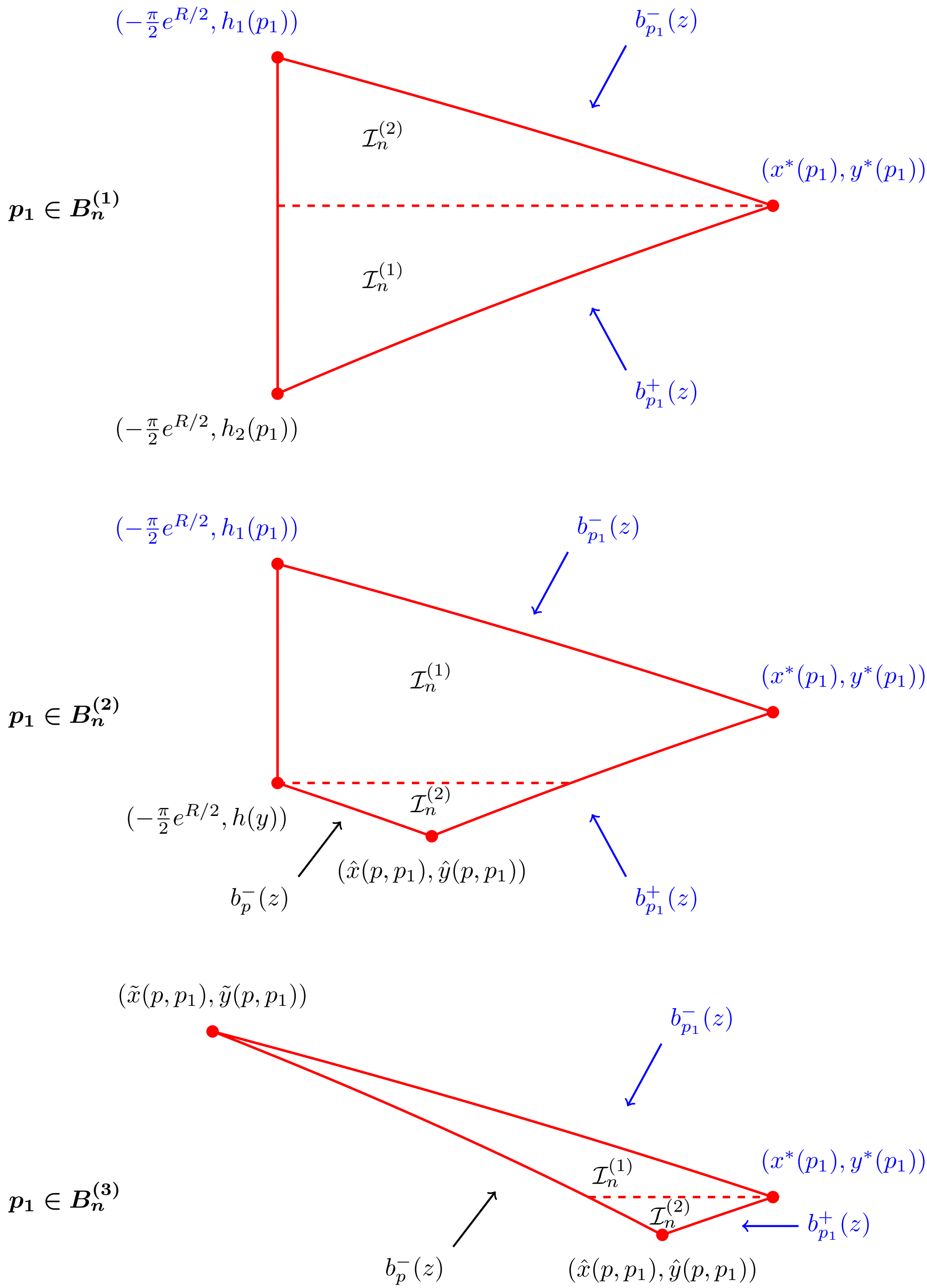

Figure 4.6: The different shapes of $\mathcal{T}(p, p_1)$ depending on the regime to which p_1 belongs. The top figure is for $p_1 \in B_n^{(1)}$, the middle for $p_1 \in B_n^{(2)}$ and the bottom one for $p_1 \in B_n^{(3)}$.

We conclude that for $p_1 \in B_n^{(1)}$:

$$\mu\left(\mathcal{T}(p,p_1)\right) = O\left(I_n^{-2\alpha} x_1 e^{\alpha y_1}\right),$$

which establishes the first part of (4.59).

Case $p_1 \in B_n^{(2)}$: $y - 2\log(I_n/(I_n - x_1)) < y_1 \leq y + 2\log(1 + \frac{x_1}{I_n})$ Here we split the integration into two parts (see Figure 4.6). Recall that $x^*(p,p_1) = x_1 - I_n$. Then, for the first part we have

$$\begin{aligned}
\mathcal{I}_n^{(1)}(p,p_1) &\leq \int_{h(y)}^{h_1(p_1)} \int_{-I_n}^{x^*(p,p_1)} f(x_2,y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2 \\
&= O\left(x_1\left(e^{-\alpha h(y)} - e^{-\alpha h_1(p_1)}\right)\right) \\
&= O\left(x_1 I_n^{-2\alpha}\left(e^{\alpha y} - e^{\alpha y_1}\left(1 + \frac{x_1}{I_n}\right)^{-2\alpha}\right)\right) \\
&= O\left(I_n^{-2\alpha} x_1 e^{\alpha y_1}\left(\left(1 - \frac{x_1}{I_n}\right)^{-2\alpha} - \left(1 + \frac{x_1}{I_n}\right)^{-2\alpha}\right)\right) \\
&= O\left(I_n^{-2\alpha} x_1 e^{\alpha y_1}\right) = O\left(I_n^{-(2\alpha-1)} e^{\alpha y}\right),
\end{aligned}$$

were we have used that $y \leq y_1 + 2\log(I_n/(I_n - x_1))$ for $p_1 \in B_n^{(2)}$ for the third line, and

$$\left(1 - \frac{x_1}{I_n}\right)^{-2\alpha} - \left(1 + \frac{x_1}{I_n}\right)^{-2\alpha} = O\left(\frac{x_1}{I_n}\right) = O\left(1\right),$$

for the last line.

For the second part we first use the upper bound on y_1 to compute that

$$\begin{aligned}
x_1 + e^{(y_1+y_2)/2} - 2I_n + e^{(y+y_2)/2} &\leq \left(e^{y/2} + e^{y_1/2}\right) e^{y_2/2} \\
&\leq e^{y/2}\left(2 + \frac{x_1}{I_n}\right) e^{y_2/2} = O\left(e^{(y+y_2)/2}\right),
\end{aligned}$$

since $|x_1| \leq I_n$. Then we have

$$\begin{aligned}
\mathcal{I}_n^{(2)}(p,p_1) &= \int_{\hat{y}(p,p_1)}^{h(y)} \int_{-e^{(y+y_2)/2}}^{x_1 + e^{(y+y_1)/2} - 2I_n} f(x_2,y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2 \\
&= O\left(e^{y/2}\int_{\hat{y}(p,p_1)}^{h(y)} e^{-(\alpha - \frac{1}{2})y_2}\,\mathrm{d}y_2\right) \\
&= O\left(e^{y/2}\left(e^{-(\alpha-\frac{1}{2})\hat{y}(p,p_1)} - e^{-(\alpha-\frac{1}{2})h(y)}\right)\right)
\end{aligned}$$

$$= O\left(e^{y/2}\left(\left(\frac{2I_n - x_1}{e^{y/2}+e^{y_1/2}}\right)^{-(2\alpha-1)} - I_n^{-(2\alpha-1)}e^{(\alpha-\frac{1}{2})y}\right)\right)$$
$$= O\left(I_n^{-(2\alpha-1)}e^{\alpha y}\right),$$

where for the last line we first used that $(2I_n - x_1)^{-(2\alpha-1)} \le I_n^{-(2\alpha-1)}$ and then

$$\left(\left(e^{y/2}+e^{y_1/2}\right)^{2\alpha-1} - e^{(\alpha-\frac{1}{2})y}\right) \le e^{(\alpha-\frac{1}{2})y}\left(\left(1+\sqrt{1+\frac{x_1}{I_n}}\right)^{2\alpha-1} - 1\right)$$
$$= O\left(e^{(\alpha-\frac{1}{2})y}\right).$$

It then follows that for $p_1 \in B_n^{(2)}$

$$\mu\left(\mathcal{T}(p,p_1)\right) = O\left(I_n^{-(2\alpha-1)}e^{\alpha y}\right).$$

Case $\boldsymbol{p_1 \in B_n^{(3)}}$: $y + 2\log(1+x_1/I_n) < y_1 \le y + 2\log(I_n/(I_n - x_1))$

$$\mathcal{I}_n^{(1)} = \int_{y^*}^{\tilde{y}}\int_{-e^{(y+y_2)/2}}^{x_1 - e^{(y_1+y_2)/2}} f(x_2,y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2$$
$$= O\left(\int_{y^*}^{\tilde{y}} x_1 e^{-\alpha y_2} - \left(e^{y_1/2} - e^{y/2}\right)e^{-(\alpha-\frac{1}{2})y_2}\,\mathrm{d}y_2\right)$$
$$= O\left(x_1\int_{y^*}^{\tilde{y}} e^{-\alpha y_2}\,\mathrm{d}y_2\right).$$

Now

$$\int_{y^*}^{\tilde{y}} e^{-\alpha y_2}\,\mathrm{d}y_2 = \frac{1}{\alpha}\left(e^{-\alpha y^*} - e^{-\alpha\tilde{y}}\right) = \frac{1}{\alpha}\left(I_n^{-2\alpha}e^{\alpha y_1} - \left(\frac{x_1}{e^{y_1/2}-e^{y/2}}\right)^{-2\alpha}\right)$$
$$= \frac{I_n^{-2\alpha}e^{\alpha y_1}}{\alpha}\left(1 - \left(1 - e^{(y-y_1)/2}\right)^{2\alpha}\left(\frac{x_1}{I_n}\right)^{-2\alpha}\right) = O\left(I_n^{-2\alpha}e^{\alpha y_1}\right),$$

and hence we have

$$\mathcal{I}_n^{(1)} = O\left(I_n^{-2\alpha}x_1 e^{\alpha y_1}\right).$$

For the second integral we have, using that $y \le y_1$ for $p_1 \in B_n^{(3)}$,

$$\mathcal{I}_n^{(2)} = \int_{\hat{y}}^{y^*}\int_{-e^{(y+y_2)/2}}^{e^{(y_1+y_2)/2}+x_1-2I_n} f(x_2,y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2$$
$$= O\left(\int_{\hat{y}}^{y^*}\left(e^{y/2}+e^{y_1/2}\right)e^{-(\alpha-\frac{1}{2})y_2}\,\mathrm{d}y_2\right)$$

$$= O\left(e^{y_1/2}\int_{\hat{y}}^{y^*} e^{-(\alpha-\frac{1}{2})y_2}\,\mathrm{d}y_2\right).$$

For the integral we have

$$\int_{\hat{y}}^{y^*} e^{-(\alpha-\frac{1}{2})y_2}\,\mathrm{d}y_2 = \frac{2}{2\alpha-1}\left(e^{-(\alpha-\frac{1}{2})\hat{y}} - e^{-(\alpha-\frac{1}{2})y^*}\right)$$
$$= \frac{2}{2\alpha-1}\left(\left(\frac{2I_n - x_1}{e^{y/2}+e^{y_1/2}}\right)^{-(2\alpha-1)} - I_n^{-(2\alpha-1)}e^{(\alpha-\frac{1}{2})y_1}\right) = O\left(I_n^{-(2\alpha-1)}e^{(\alpha-\frac{1}{2})y_1}\right),$$

where we have used the upper bound on y_1 and the fact that $2I_n - x_1 = \Theta(I_n)$ for all $x_1 \in [-I_n, I_n]$. We conclude that

$$\mathcal{I}_n^{(2)} = O\left(I_n^{-(2\alpha-1)}x_1 e^{\alpha y}\right)$$

and hence for $p_1 \in B_n^{(3)}$

$$\mu(\mathcal{T}(p,p_1)) = O\left(I_n^{-2\alpha}x_1 e^{\alpha y}\right) = O\left(I_n^{-(2\alpha-1)}e^{\alpha y}\right).$$

Integration $\mu(\mathcal{T}(p,p_1))$ with respect to p_1

We now proceed with the second part of the computation leading to (4.56). Here we will integrate $\mu(\mathcal{T}(p,p'))(p,p_1)$ over the region $B_n := B_n^{(1)} \cup B_n^{(2)} \cup B_n^{(3)}$, see Figure 4.5. Let us first identify the boundaries of these areas.

The area $B_n^{(1)}$ is bounded from above by the line given by the equation

$$y_1 = y - 2\log\left(\frac{I_n}{I_n - x_1}\right).$$

Solving this for x_1 yields $x_1 = I_n\left(1 - e^{(y_1-y)/2}\right)$ and hence the area $B_n^{(1)}$ is given by

$$B_n^{(1)} = \left\{(x_1,y_1) \,:\, 0 \le y_1 \le y, \quad 0 \le x_1 \le I_n\left(1 - e^{(y_1-y)/2}\right) \wedge e^{(y+y_1)/2}\right\}.$$

In a similar way we have that $B_n^{(2)}$ is bounded from above by the line

$$y_1 = y + 2\log\left(\frac{I_n}{I_n + x_1}\right),$$

which yields $x_1 = I_n\left(e^{(y_1-y)/2} - 1\right)$. The lower red boundary is the upper boundary of $B_n^{(2)}$ and hence we have

$$B_n^{(2)} = \{(x_1,y_1) \,:\, h_*(y) \le y_1 \le h^*(y),$$

$$I_n\left(1-e^{(y_1-y)/2}\right) \vee I_n\left(e^{(y_1-y)/2}-1\right) \le x_1 \le e^{(y+y_1)/2}\Big\}.$$

We continue in the same way for $B_n^{(3)}$

$$\begin{aligned} B_n^{(3)} = \{(x_1, y_1) \,:\, & y \le y_1 \le R, \\ & I_n\left(1-e^{(y-y_1)/2}\right) \le x_1 \le I_n\left(e^{(y_1-y)/2}-1\right) \wedge e^{(y+y_1)/2} \wedge I_n\Big\}. \end{aligned}$$

With these characterizations of the areas we now integrate $\mu(\mathcal{T}(p,p_1))$ over B_n, splitting the computations over the three different areas.

Integration over $B_n^{(1)}$: We use that

$$I_n\left(1-e^{(y_1-y)/2}\right) \wedge e^{(y+y_1)/2} \le I_n\left(1-e^{(y_1-y)/2}\right),$$

so that

$$\begin{aligned} &\int_{B_n^{(1)}} \mu\left(\mathcal{T}(p,p_1)\right) f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \\ &\quad\le \int_0^y \int_0^{I_n(1-e^{(y_1-y)/2})} \mu\left(\mathcal{T}(p,p_1)\right) f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \\ &\quad= O\left(I_n^{-2\alpha}\int_0^y\int_0^{e^{(y+y_1)/2}} x_1\,\mathrm{d}x_1\,\mathrm{d}y_1\right) \\ &\quad= O\left(I_n^{-(2\alpha-1)}\int_0^y \left(1-e^{(y_1-y)/2}\right)^2 \mathrm{d}y_1\right) \\ &\quad= O\left(I_n^{-(2\alpha-1)}y\right) = O\left(yn^{-(2\alpha-1)}\right). \end{aligned}$$

Integration over $B_n^{(2)}$: We will show that

$$\mu(B_n^{(2)}) = O\left(I_n^{-1}e^{(2-\alpha)y}\right), \tag{4.60}$$

which together with (4.59) yields

$$\begin{aligned} \int_{B_n^{(2)}} \mu\left(\mathcal{T}(p,p_1)\right) f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 &= O\left(\mu(B_n^{(2)})I_n^{-(2\alpha-1)}e^{\alpha y}\right) \\ &= O\left(I_n^{-2\alpha}e^{2y}\right). \end{aligned}$$

The integration is split into two parts that are determined by $I_n\left(1-e^{(y_1-y)/2}\right) \vee I_n\left(e^{(y_1-y)/2}-1\right)$:

$$\mu(B_n^{(3)}) = \int_{h_*(y)}^y \int_{I_n(1-e^{(y_1-y)/2})}^{e^{(y+y_1)/2}} f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1$$

$$+\int_y^{h^*(y)} \int_{I_n(e^{(y_1-y)/2}-1)}^{e^{(y+y_1)/2}} f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1.$$

For the first integral we use that $e^{(y+y_1)/2} - I_n(1-e^{(y_1-y)/2}) \le e^{y_1/2}\left(e^{y/2}+e^{-y/2}\right)$ to obtain

$$\begin{aligned}
&\int_{h_*(y)}^{y} \int_{I_n(1-e^{(y_1-y)/2})}^{e^{(y+y_1)/2}} f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \\
&\quad = O\left(e^{y/2}\int_{h_*(y)}^{y} e^{-(\alpha-\frac{1}{2})y_1}\,\mathrm{d}y_1\right) \\
&\quad = O\left(e^{y/2}\left(e^{-(\alpha-\frac{1}{2})y} - e^{-(\alpha-\frac{1}{2})y}\left(\frac{I_n}{I_n+e^y}\right)^{-(2\alpha-1)}\right)\right) \\
&\quad = O\left(I_n^{-1}e^{(2-\alpha)y}\right).
\end{aligned}$$

For the second integral note that $e^{(y+y_1)/2} - I_n(e^{(y_1-y)/2}-1) \le e^{(y+y_1)/2}$ and hence

$$\begin{aligned}
&\int_{y}^{h^*(y)} \int_{I_n(e^{(y_1-y)/2}-1)}^{e^{(y+y_1)/2}} f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \\
&\quad = O\left(e^{y/2}\int_{y}^{h^*(y)} e^{-(\alpha-\frac{1}{2})y_1}\,\mathrm{d}y_1\right) \\
&\quad = O\left(e^{y/2}\left(e^{-(\alpha-\frac{1}{2})y} - e^{-(\alpha-\frac{1}{2})y}\left(\frac{I_n}{I_n-e^y}\right)^{-(2\alpha-1)}\right)\right) \\
&\quad = O\left(I_n^{-1}e^{(2-\alpha)y}\right),
\end{aligned}$$

so that (4.60) follows.

Integration over $B_n^{(3)}$: For this case we show that

$$\mu(B_n^{(3)}) = O\left(e^{(1-\alpha)y}\right), \tag{4.61}$$

so that

$$\begin{aligned}
\int_{B_n^{(3)}} \mu\left(\mathcal{T}(p,p_1)\right) f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 &= O\left(\mu(B_n^{(2)})I_n^{-(2\alpha-1)}e^{\alpha y}\right) \\
&= O\left(I_n^{-(2\alpha-1)}e^{y}\right).
\end{aligned}$$

Here the integral is split into three parts:

$$\mu(B_n^{(3)}) = \int_y^{h^*(y)} \int_{I_n(1-e^{(y-y_1)/2})}^{I_n(e^{(y_1-y)/2}-1)} f(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1$$

$$
\begin{aligned}
&+ \int_{h^*(y)}^{h(y)} \int_{I_n(1-e^{(y-y_1)/2})}^{e^{(y+y_1)/2}} f(x_1, y_1) \, \mathrm{d}x_1 \, \mathrm{d}y_1 \\
&+ \int_{h(y)}^{R} \int_{I_n(1-e^{(y-y_1)/2})}^{I_n} f(x_1, y_1) \, \mathrm{d}x_1 \, \mathrm{d}y_1.
\end{aligned}
$$

Let us first focus on the first integral. Since $I_n(e^{(y_1-y)/2}-1)-I_n(1-e^{(y-y_1)/2}) \leq I_n e^{(y_1-y)/2}$ we get, using similar arguments as above

$$
\begin{aligned}
\int_y^{h^*(y)} \int_{I_n(1-e^{(y-y_1)/2})}^{I_n(e^{(y_1-y)/2}-1)} f(x_1, y_1) \, \mathrm{d}x_1 \, \mathrm{d}y_1 &= O\left(I_n e^{-y/2} \int_y^{h^*(y)} e^{-(\alpha - \frac{1}{2})y_1} \, \mathrm{d}y_1 \right) \\
&= O\left(I_n e^{-\alpha y} \left(1 - \left(\frac{I_n}{I_n - e^y} \right)^{-(2\alpha-1)} \right) \right) \\
&= O\left(e^{(1-\alpha)y} \right).
\end{aligned}
$$

Proceeding to the second integral, we first note that $e^{(y+y_1)/2}-I_n(1-e^{(y-y_1)/2}) = O\left(I_n e^{(y_1-y)/2}\right)$ so that similar calculations as before yield

$$
\begin{aligned}
\int_{h^*(y)}^{h(y)} \int_{I_n(1-e^{(y-y_1)/2})}^{e^{(y+y_1)/2}} f(x_1, y_1) \, \mathrm{d}x_1 \, \mathrm{d}y_1 &= O\left(I_n e^{-y/2} \int_{h^*(y)}^{h(y)} e^{-(\alpha - \frac{1}{2})y_1} \, \mathrm{d}y_1 \right) \\
&= O\left(e^{(1-\alpha)y} \right).
\end{aligned}
$$

□

4.6 Concentration for $c(k_n; G_{\mathrm{box}})$ (Proving Proposition 4.3.5)

In this section we establish a concentration result for the local clustering function $c^*(k; G_{\mathrm{box}})$ in the finite box model G_{box}. Similar to the previous section we will focus on typical points $p = (0, y)$ with $y \in \mathcal{K}_C(k_n)$.

4.6.1 The main contribution of triangles

Recall that $N_{\mathrm{box}}(k_n)$ denotes the number of vertices in G_{box} with degree k_n. We first write

$$
c^*(k_n; G_{\mathrm{box}}) = \frac{T_{\mathrm{box}}(k_n)}{\binom{k_n}{2} \mathbb{E}\left[N_{\mathrm{box}}(k_n)\right]},
$$

where

$$
T_{\mathrm{box}}(k_n) = \sum_{p \in \mathcal{P}} \mathbb{1}_{\{\deg_{\mathrm{box}}(p) = k_n\}} \sum_{(p_1, p_2) \in \mathcal{P} \setminus \{p\}}^{\neq} \mathbb{1}_{\{p_1 \in \mathcal{B}_{\mathrm{box}}(p)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{box}}(p)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{box}}(p_1)\}}.
$$

In particular, the variance of $c^*(k_n; G_{\mathrm{box}})$ is determined by the variance of $T_{\mathrm{box}}(k_n)$.

Next, recall the adjusted triangle count function

$$\widetilde{T}_{\mathrm{box}}(p_0) = \sum_{(p_1,p_2)\in\mathcal{P}\setminus\{p_0\}}^{\neq} \widetilde{T}_{\mathrm{box}}(p_0,p_1,p_2),$$

where

$$\widetilde{T}_{\mathrm{box}}(p_0,p_1,p_2) = \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{box}}(p_0)\}}\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{box}}(p_0)\}}\mathbb{1}_{\{p_2\in\mathcal{B}_{\infty}(p_1)\cap\mathcal{R}\}},$$

as well as the definition of $\mathcal{K}_C(k_n)$

$$\mathcal{K}_C(k_n) = \left\{y \in \mathbb{R}_+ : \frac{k_n - C\sqrt{k_n \log(k_n)}}{\xi} \vee 1 \le e^{\frac{y}{2}} \le \frac{k_n + C\sqrt{k_n \log(k_n)}}{\xi}\right\},$$

and write $\mathcal{R}(k_n, C) = [-I_n, I_n] \times \mathcal{K}_C(k_n)$ for the part of the box $\mathcal{R}$ with heights in $\mathcal{K}_C(k_n)$. Slightly abusing notation, we will define the corresponding triangle degree function

$$\widetilde{T}_{\mathrm{box}}(k_n, C) = \sum_{p\in\mathcal{P}\cap\mathcal{R}(k_n,C)} \mathbb{1}_{\{\deg_{\mathrm{box}}(p)=k_n\}}\widetilde{T}_{\mathrm{box}}(p), \tag{4.62}$$

and with that a different clustering function

$$\widetilde{c}_{\mathrm{box}}(k_n) = \frac{\widetilde{T}_{\mathrm{box}}(k_n, C)}{\binom{k_n}{2}\mathbb{E}\left[N_{\mathrm{box}}(k_n)\right]}. \tag{4.63}$$

The idea is that the main contribution of triangles of degree k_n to the triangle count $T_{\mathrm{box}}(k_n)$ is given by $\widetilde{T}_{\mathrm{box}}(k_n, C)$. Therefore, in order to prove Proposition 4.3.5 it suffices to show that $\widetilde{T}_{\mathrm{box}}(k_n, C)$ is sufficiently concentrated around its mean. This last part is done in the following proposition:

Proposition 4.6.1 (Concentration $\widetilde{T}_{\mathrm{box}}(k_n, C)$)**.** *Let $\alpha > \frac{1}{2}$, $\nu > 0$ and let $(k_n)_{n\ge1}$ be any positive sequence satisfying $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$. Then for any $C > 0$, as $n \to \infty$,*

$$\mathbb{E}\left[\widetilde{T}_{box}(k_n, C)^2\right] = (1 + o(1))\,\mathbb{E}\left[\widetilde{T}_{box}(k_n, C)\right]^2.$$

We first use this result to prove Proposition 4.3.5. The remainder of this section is devoted to the proof of Proposition 4.6.1. The final proof can be found in Section 4.6.3.

Proof of Proposition 4.3.5. We bound the expectation as

$$\mathbb{E}\left[\left|c^*(k_n; G_{\mathrm{box}}) - \mathbb{E}\left[c^*(k_n; G_{\mathrm{box}})\right]\right|\right] \le \frac{\mathbb{E}\left[\left|\widetilde{T}_{\mathrm{box}}(k_n, C) - \mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(k_n, C)\right]\right|\right]}{\binom{k_n}{2}\mathbb{E}\left[N_{\mathrm{box}}(k_n)\right]}$$

$$+ 2\mathbb{E}\left[|c^*(k_n; G_{\text{box}}) - \widetilde{c}_{\text{box}}(k_n)|\right].$$

We will show that both terms are $o(s(k_n))$.

First we note that $\mathbb{1}_{\{p_2 \in \mathcal{B}_\infty(p_1) \cap \mathcal{R}\}} \leq \mathbb{1}_{\{p_2 \in \mathcal{B}_{\text{box}}(p_1)\}}$ and hence $\widetilde{T}_{\text{box}}(p) \leq T_{\text{box}}(p)$. This implies that

$$\widetilde{c}_{\text{box}}(k_n) = \frac{\widetilde{T}_{\text{box}}(k_n, C)}{\binom{k_n}{2}\mathbb{E}\left[N_{\text{box}}(k_n)\right]} \leq c^*(k_n; G_{\text{box}}).$$

and therefore

$$\mathbb{E}\left[|c^*(k_n; G_{\text{box}}) - \widetilde{c}_{\text{box}}(k_n)|\right] = \mathbb{E}\left[c^*(k_n; G_{\text{box}})\right] - \mathbb{E}\left[\widetilde{c}_{\text{box}}(k_n)\right].$$

For the expectation of $\widetilde{T}_{\text{box}}(k_n, C)$ we use that

$$\mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\middle| \deg_{\text{box}}(p) = k_n\right] = \binom{k_n}{2}\mu\left(\mathcal{B}_{\text{box}}\left(y\right)\right)^{-2}\mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\right],$$

to get

$$\begin{aligned}
&\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right] \\
&= \int_{\mathcal{R}(k_n, C)} \mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\middle| \deg_{\text{box}}(p) = k_n\right] \rho_{\text{box}}(y, k_n) f(x, y) \, \mathrm{d}x \, \mathrm{d}y \\
&= (1 + o(1))\binom{k_n}{2} \int_{\mathcal{R}(k_n, C)} \mu\left(\mathcal{B}_{\text{box}}\left(y\right)\right)^{-2} \mathbb{E}\left[\widetilde{T}_{\text{box}}(y)\right] \rho_{\text{box}}(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&= (1 + o(1))\frac{1}{2} \int_{\mathcal{R}(k_n, C)} \mathbb{E}\left[\widetilde{T}_{\text{box}}(y)\right] \rho_{\text{box}}(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y \\
&= (1 + o(1)) n \binom{k_n}{2} \int_0^\infty P(y) \rho(y, k_n) \alpha e^{-\alpha y} \, \mathrm{d}y,
\end{aligned}$$

where the last line is due to Corollary 4.5.4. In particular, since the last integral is $\Theta\left(k_n^{-(2\alpha+1)} s(k_n)\right)$, we conclude that

$$\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right] = \Theta\left(n k_n^{-(2\alpha-1)} s(k_n)\right). \tag{4.64}$$

Since $\mathbb{E}\left[N_{\text{box}}(k_n)\right] = (1 + o(1)\, n p_{k_n}$, it follows that

$$\widetilde{c}_{\text{box}}(k_n) = \frac{\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]}{\binom{k_n}{2}\mathbb{E}\left[N_{\text{box}}(k_n)\right]} = (1 + o(1))\frac{\int_0^\infty P(y) \alpha e^{-\alpha y} \, \mathrm{d}y}{p_{k_n}} = (1 + o(1))\gamma(k_n).$$

On the other hand, Proposition 4.3.6 implies that $\mathbb{E}\left[c^*(k_n; G_{\text{box}})\right] = (1+o(1))\gamma(k_n)$ and thus we conclude that

$$2\mathbb{E}\left[|c^*(k_n; G_{\text{box}}) - \widetilde{c}_{\text{box}}(k_n)|\right] = o\left(\gamma(k_n)\right) = o\left(s(k_n)\right).$$

For the remaining term we use Hölder's inequality and Proposition 4.6.1 to obtain that

$$\begin{aligned}\mathbb{E}\left[\left|\widetilde{T}_{\text{box}}(k_n, C) - \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]\right|\right] &\leq \left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)^2\right] - \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]^2\right)^{\frac{1}{2}} \\ &= o\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]\right).\end{aligned}$$

This implies

$$\frac{\mathbb{E}\left[\left|\widetilde{T}_{\text{box}}(k_n, C) - \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]\right|\right]}{\binom{k_n}{2}\mathbb{E}\left[N_{\text{box}}(k_n)\right]} = o\left(\frac{\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]}{\binom{k_n}{2}\mathbb{E}\left[N_{\text{box}}(k_n)\right]}\right) = o\left(s(k_n)\right),$$

which finishes the proof. □

We note that the above proof establishes the following important result:

Corollary 4.6.2. *Let $k_n \to \infty$. Then, as $n \to \infty$,*

$$\mathbb{E}\left[\left|c^*(k_n; G_{\text{box}}) - \widetilde{c}_{\text{box}}(k_n)\right|\right] = o\left(s(k_n)\right).$$

4.6.2 Joint neighbourhoods and degrees in G_{box}

To prove Proposition 4.6.1 we need to understand the joint degree distribution in G_{box}. This subsequently requires us to analyse the joint neighbourhoods in G_{box} of two points $p, p' \in \mathcal{R}$. We start with a general result for near-independent Poisson random variables.

Lemma 4.6.3. *Let $k_n \to \infty$ and $X_1 = \text{Po}(\lambda_1(n))$, $X_2 = \text{Po}(\lambda_2(n))$ and $Y = \text{Po}(\lambda_3(n))$, be three Poisson random variables where $\lambda_3(n) = O\left(k_n^{1-\varepsilon}\right)$, for some $0 < \varepsilon < 1$ and for some $C > 0$,*

$$k_n - C\sqrt{k_n \log(k_n)} \leq \lambda_i(n) + \lambda_3(n) \leq k_n + C\sqrt{k_n \log(k_n)},$$

for $i = 1, 2$. Then, as $n \to \infty$

$$\mathbb{P}\left(X_1 + Y = k_n, X_2 + Y = k_n\right) = (1 + o(1))\mathbb{P}\left(X_1 + Y = k_n\right)\mathbb{P}\left(X_2 + Y = k_n\right).$$

Proof. First we write

$$\mathbb{P}\left(X_1 + Y = k_n, X_2 + Y = k_n\right) = \sum_{t=0}^{\infty} \mathbb{P}\left(X_1 = k_n - t\right)\mathbb{P}\left(X_2 = k_n - t\right)\mathbb{P}\left(Y = t\right).$$

Now fix a $C_1 > 0$ and define the set

$$A_n := \left\{t \in \mathbb{R}_+ \,:\, \lambda_3(n) - C_1\sqrt{k_n^{1-\varepsilon}\log(k_n)} \leq t \leq \lambda_3(n) + C_1\sqrt{k_n^{1-\varepsilon}\log(k_n)}\right\}.$$

Then by a Chernoff bound (c.f. (1.12))

$$\begin{aligned}
&\sum_{t\in\mathbb{R}_+\setminus A_n} \mathbb{P}\left(X_1 = k_n - t\right)\mathbb{P}\left(X_2 = k_n - t\right)\mathbb{P}\left(Y = t\right)\\
&\quad \le \mathbb{P}\left(Y > \lambda_3(n) + C_1\sqrt{k_n^{1-\varepsilon}\log(k_n)}\right) + \mathbb{P}\left(Y < \lambda_3(n) - C_1\sqrt{k_n^{1-\varepsilon}\log(k_n)}\right)\\
&\quad = \mathbb{P}\left(|\mathrm{Po}(\lambda_3(n)) - \lambda_3(n)| > C_1\sqrt{k_n^{1-\varepsilon}\log(k_n)}\right) \le 2k_n^{-\frac{C_1}{4}},
\end{aligned}$$

and hence

$$\begin{aligned}
&\mathbb{P}\left(X_1 + Y = k_n, X_2 + Y = k_n\right)\\
&= \sum_{t\in A_n} \mathbb{P}\left(X_1 = k_n - t\right)\mathbb{P}\left(X_2 = k_n - t\right)\mathbb{P}\left(Y = t\right) + O\left(k_n^{-(1+C_1^2)/2}\right).
\end{aligned}$$

Next, for $i = 1, 2$ we have by assumption on $\lambda_i(n) + \lambda_3(n)$ that

$$\begin{aligned}
\mathbb{P}\left(X_i + Y = k_n\right) &\ge \frac{\left(k_n + C\sqrt{k_n\log(k_n)}\right)^{k_n}}{k_n!} e^{-\left(k_n + C\sqrt{k_n\log(k_n)}\right)}\\
&\ge e^{-1}k_n^{-1/2}\left(1 + C\sqrt{\frac{\log(k_n)}{k_n}}\right)^{k_n} e^{-C\sqrt{k_n\log(k_n)}}\\
&\ge e^{-1}k_n^{-1/2}e^{k_n\log\left(1+C\sqrt{\frac{\log(k_n)}{k_n}}\right) - C\sqrt{k_n\log(k_n)}}\\
&\ge e^{-1}k_n^{-1/2}e^{-\frac{C^2}{2}\log(k_n)} = e^{-1}k_n^{-\frac{1+C^2}{2}}.
\end{aligned}$$

where we have also used that $\log(1+x) \ge x - x^2/2$, for $0 \le x \le 1$ and $k_n! \ge e\sqrt{k_n}k_n^{k_n}e^{-k_n}$. Hence, by taking $C_1 > 4(1+C^2)$ we get that

$$k_n^{-\frac{C_1}{4}} = o\left(\mathbb{P}\left(X_1 + Y = k_n\right)\mathbb{P}\left(X_2 + Y = k_n\right)\right).$$

It remains to show that

$$\begin{aligned}
&\sum_{t\in A_n} \mathbb{P}\left(X_1 = k_n - t\right)\mathbb{P}\left(X_2 = k_n - t\right)\mathbb{P}\left(Y = t\right)\\
&\quad = (1 + o(1))\mathbb{P}\left(X_1 + Y = k_n\right)\mathbb{P}\left(X_2 + Y = k_n\right).
\end{aligned}$$

For this take any $s \in A_n$ so that $|t - s| \le 2C_1\sqrt{k_n^{1-\varepsilon}\log(k_n)}$ and note that there exists a δ_n satisfying $|\delta_n| \le 2C\sqrt{k_n\log(k_n)}$, for n large enough, such that $k_n - t = \lambda_1(n) + \delta_n$. It then follows that, uniformly in t, s and δ_n, as $n \to \infty$,

$$\frac{\mathbb{P}\left(X_2 = k_n - t\right)}{\mathbb{P}\left(X_2 = k_n - s\right)} = \frac{\mathbb{P}\left(X_2 = k_n - t\right)}{\mathbb{P}\left(X_2 = k_n - t - (s - t)\right)}$$

$$
\begin{aligned}
&= \frac{(k_n - t - (s-t))!}{(k_n - t)!}\lambda_1(n)^{s-t} \\
&\sim (k_n - t - (s-t))^{-(s-t)}\lambda_1(n)^{s-t} \\
&= (\lambda_1(n) + \delta_n - (s-t))^{-(s-t)}\,\lambda_1(n)^{s-t} \\
&= \left(1 + \frac{\delta_n - (s-t)}{\lambda_1(n)}\right)^{s-t} \\
&\sim e^{\frac{(s-t)\delta_n}{\lambda_1(n)}} e^{-\frac{(s-t)^2}{\lambda_1(n)}} \sim 1,
\end{aligned}
$$

where the last line follows since both $\frac{(s-t)\delta_n}{\lambda_1(n)} \to 0$ and $\frac{(s-t)^2}{\lambda_1(n)} \to 0$ as $n \to \infty$. In particular,

$$\mathbb{P}(X_2 = k_n - t) = (1 + o(1))\mathbb{P}(X_2 = k_n - s),$$

uniformly for all $t, s \in A_n$ and therefore, since

$$1 = \sum_{s=0}^{\infty} \mathbb{P}(Y = s) = (1 + o(1)) \sum_{s \in A_n} \mathbb{P}(Y = s),$$

we conclude that

$$
\begin{aligned}
&\sum_{t \in A_n} \mathbb{P}(X_1 = k_n - t)\,\mathbb{P}(X_2 = k_n - t)\,\mathbb{P}(Y = t) \\
&\quad = (1 + o(1)) \sum_{t \in A_n} \mathbb{P}(X_1 = k_n - t)\,\mathbb{P}(X_2 = k_n - t)\,\mathbb{P}(Y = t) \sum_{s \in A_n} \mathbb{P}(Y = s) \\
&\quad = (1 + o(1)) \sum_{t \in A_n} \mathbb{P}(X_1 = k_n - t)\,\mathbb{P}(Y = t) \sum_{s \in A_n} \mathbb{P}(X_2 = k_n - s)\,\mathbb{P}(Y = s) \\
&\quad = (1 + o(1))\mathbb{P}(X_1 + Y = k_n)\,\mathbb{P}(X_2 + Y = k_n),
\end{aligned}
$$

from which the result follows. □

To see how this lemma can be applied to analyze the joint degree distribution in G_{box}, fix two points $p, p' \in \mathcal{R}$ and denote by

$$\rho_{\text{box}}(p, p', k, k') := \mathbb{P}(\text{Po}(\mu(\mathcal{B}_{\text{box}}(p))) = k, \text{Po}(\mu(\mathcal{B}_{\text{box}}(p'))) = k') \tag{4.65}$$

the joint degree distribution. Then if we define,

$$
\begin{aligned}
X_1(p, p') &:= \text{Po}(\mu(\mathcal{B}_{\text{box}}(p) \setminus \mathcal{B}_{\text{box}}(p'))), \\
X_2(p, p') &:= \text{Po}(\mu(\mathcal{B}_{\text{box}}(p') \setminus \mathcal{B}_{\text{box}}(p))), \\
Y(p, p') &:= \text{Po}(\mu(\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p'))),
\end{aligned}
$$

it follows that

$$\rho_{\text{box}}(p, p', k_n, k_n) = \mathbb{P}(X_1(p, p') + Y(p, p') = k_n, X_2(p, p') + Y(p, p') = k_n).$$

Now, if $y, y' \in \mathcal{K}_C(k_n)$ the three Poisson random variables defined above satisfy the condition of Lemma 4.6.3 regarding the sum $\lambda_i(n) + \lambda_3(n)$. Therefore, if in addition $\mu\left(\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')\right) = O\left(k_n^{1-\varepsilon}\right)$, for some $0 < \varepsilon < 1$, we have that

$$\rho_{\text{box}}(p, p', k_n, k_n) = (1 + o(1))\rho_{\text{box}}(p, k_n)\rho_{\text{box}}(p', k_n).$$

To make this more precise, we define, for any $0 < \varepsilon < 1$, the set

$$\mathcal{E}_\varepsilon(k_n) = \left\{(p, p') \in \mathcal{R} \times \mathcal{R} \,:\, y, y' \in \mathcal{K}_C(k_n) \text{ and } |x - x'|_n > k_n^{1+\varepsilon}\right\}, \tag{4.66}$$

where $|x|_n = \min\{|x|, \pi e^{R/2} - |x|\}$ denotes the norm on the finite box $\mathcal{R}$ where the left and right boundaries are identified. We will show (see Corollary 4.6.6) that for all $(p, p') \in \mathcal{E}_\varepsilon(k_n)$ it holds that $\mu\left(\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')\right) = O\left(k_n^{1-\varepsilon}\right)$ and hence the joint degree distribution factorizes on this set. We will use this set later in Section 4.6.3 to prove Proposition 4.6.1. The main idea behind the above result is that if p and p' are sufficiently separated in the x-direction, then the overlap of their neighbourhoods $\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')$ is of smaller order than $\mu\left(\mathcal{B}_{\text{box}}(p)\right) + \mu\left(\mathcal{B}_{\text{box}}(p')\right)$. We shall therefore proceed with analyzing the joint neighbourhoods in G_{box}.

Common neighbourhoods

Let $p, p' \in \mathcal{R}$ and denote by $\mathcal{N}_{\text{box}}(p, p')$ the number of common neighbours of p and p'. We shall establish an upper bound on the expected number of joint neighbours when p and p' are sufficiently separated. Observe that $\mathbb{E}\left[\mathcal{N}_{\text{box}}(p, p')\right] = \mu\left(\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')\right)$.

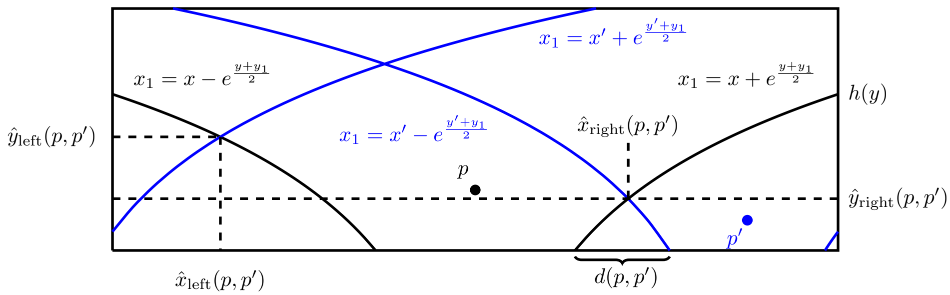

Figure 4.7: Schematic representation of the neighbourhoods of p and p' in G_{box} when $|x - x'| > e^{\frac{y}{2}} + e^{\frac{y'}{2}}$ used for the proof of Lemma 4.6.4. Note that although here $p' \notin \mathcal{B}_{\text{box}}(p)$, this is not true in general. This situation was merely chosen to improve readability of the figure.

We start by analyzing the shape of the joint neighbourhood. Due to symmetry and the fact that we have identified the left and right boundaries of the box $\mathcal{R}$, we can, without loss of generality, assume that $p = (0, y)$ and $p' = (x', y')$ with $x' > 0$.

To understand the computation it is helpful to have a picture. Figure 4.7 shows such an example. There are several different quantities that are important. The first ones are the heights where the left and right boundaries of the ball $\mathcal{B}_{\text{box}}(p)$ hit the boundaries of the box $\mathcal{R}$. Since $x = 0$ these heights are the same and we denote their common value by $h(y)$. We also need to know the coordinates $\hat{y}_{\text{right}}(p, p')$ and $\hat{x}_{\text{right}}(p, p')$ of the intersection of the right boundary of the neighbourhood of p with the left boundary of the neighbourhood of p' and those for the intersection of the left boundary of the neighbourhood of p with the right boundary of the neighbourhood of p', which we denote by $\hat{y}_{\text{left}}(p, p')$ and $\hat{x}_{\text{left}}(p, p')$. Finally we will denote by $d(p, p')$ the distance between the lower right boundary of $\mathcal{B}_{\text{box}}(p)$ and the lower left of $\mathcal{B}_{\text{box}}(p')$, which is positive only when the bottom parts of both neighbourhoods do not intersect, as is the case in Figure 4.7. The condition $d(p, p') > 0$ is exactly the right notion for p and p' being sufficiently separated.

Note that $\hat{y}_{\text{left}}(p, p')$ and $\hat{x}_{\text{left}}(p, p')$ correspond to, respectively, $\hat{y}(p, p')$ and $\hat{x}(p, p')$ considered in Section 4.5.2. The derivation of $\hat{y}_{\text{right}}(p, p')$ and $\hat{x}_{\text{right}}(p, p')$ is done in a similar manner and we omit the details here. The full expressions of all these functions are given below for further reference:

$$h(y) = R - y + 2\log\left(\frac{\pi}{2}\right), \tag{4.67}$$

$$h_1(p') = 2\log\left(x' + \frac{\pi}{2}e^{\frac{R}{2}}\right) - y', \tag{4.68}$$

$$h_2(p') = 2\log\left(\frac{\pi}{2}e^{\frac{R}{2}} - x'\right) - y', \tag{4.69}$$

$$\hat{y}_{\text{right}}(p, p') = 2\log\left(\frac{x'}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right), \tag{4.70}$$

$$\hat{x}_{\text{right}}(p, p') = \frac{x'}{1 + e^{\frac{y'-y}{2}}}, \tag{4.71}$$

$$\hat{y}_{\text{left}}(p, p') = 2\log\left(\frac{\pi e^{R/2} - x'}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right), \tag{4.72}$$

$$\hat{x}_{\text{left}}(p, p') = \frac{x' - \pi e^{R/2}}{1 + e^{\frac{y'-y}{2}}}, \tag{4.73}$$

$$d(p, p') = |x - x'|_n - \left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right). \tag{4.74}$$

The following result shows that if $d(p, p') > 0$, then the expected number of common neighbours is $o\left(\mu\left(\mathcal{B}_{\text{box}}(p)\right) + \mu\left(\mathcal{B}_{\text{box}}(p')\right)\right)$:

Lemma 4.6.4. *Let $p, p' \in \mathcal{R}$. Then, whenever $|x - x'|_n > \left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)$,*

$$\mathbb{E}\left[\mathcal{N}_{box}(p, p')\right] \le \mu\left(\mathcal{B}_{box}(p)\right)\left(\left(\frac{|x - x'|}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right)^{-(2\alpha-1)} + \frac{\nu}{\xi}e^{-(\alpha-\frac{1}{2})(R-y)}\right).$$

Proof. Again, without loss of generality, we assume that $p = p_0 = (0, y)$ and $p' = (x', y')$ with $0 \le x' \le \frac{\pi}{2} e^{R/2}$. Note that since $0 < x' \le \frac{\pi}{2} e^{R/2}$, it holds that $\hat{y}_{\text{right}}(p, p') \le \hat{y}_{\text{left}}(p, p')$. We write $\hat{y}$ for $\hat{y}_{\text{right}}(p, p')$ and observe that below $\hat{y}$ the balls $\mathcal{B}_{\text{box}}(p)$ and $\mathcal{B}_{\text{box}}(p')$ are disjoint. Therefore, if we define

$$A := \{p_1 = (x_1, y_1) \in \mathcal{R} \cap \mathcal{B}_{\text{box}}(p) \,:\, y_1 \ge \hat{y}\},$$

then

$$\mathbb{E}\left[\mathcal{N}_{\text{box}}(p, p')\right] \le \mu(A).$$

We proceed with computing the right-hand side as

$$\begin{aligned}
\mu(A) &= \int_{\hat{y}}^{h(y)} \int_{-e^{\frac{y+y_1}{2}}}^{e^{\frac{y'+y_1}{2}}} f(x_1, y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 + \int_{h(y)}^{R} \int_{-\frac{\pi}{2}e^{R/2}}^{\frac{\pi}{2}e^{R/2}} f(x_1, y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \\
&= \frac{2\alpha\nu}{\pi} e^{\frac{y}{2}} \int_{\hat{y}}^{h(y)} e^{-(\alpha - \frac{1}{2})y_1}\,\mathrm{d}y_1 + \alpha\nu e^{R/2} \int_{h(y)}^{R} e^{-\alpha y_1}\,\mathrm{d}y_1 \\
&\le \xi \left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right) e^{-(\alpha - \frac{1}{2})\hat{y}} + \nu e^{R/2} e^{-\alpha h(y)} \\
&= \mu\left(\mathcal{B}_{\text{box}}(p)\right) \left(e^{-(\alpha - \frac{1}{2})\hat{y}} + \frac{\nu}{\xi} e^{-(\alpha - \frac{1}{2})(R - y)}\right).
\end{aligned}$$

The result follows by plugging in

$$\hat{y} := \hat{y}_{\text{right}}(p, p') = 2 \log\left(\frac{x'}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right),$$

and noting that x' is the same as $|x - x'|$, by our generalization step.

□

We also prove a similar result for the Poissonized KPKVB model G_{Po}:

Lemma 4.6.5. *Let $0 < \varepsilon < 1$, $p, p' \in \mathcal{R}$ with $y, y' \le (1 - \varepsilon)R$ and denote by $\mathcal{N}_{\text{Po}}(p, p')$ the number of joint neighbours of p, p' in G_{Po}. Then, whenever $|x - x'|_n > \left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)\left(1 + \frac{\pi^2 K}{4}\right)$,*

$$\mathbb{E}\left[\mathcal{N}_{\text{Po}}(p, p')\right] \le \mu\left(\mathcal{B}(p)\right) \left(e^{(2\alpha - 1)\lambda} \left(\frac{|x - x'|}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right)^{-(2\alpha - 1)} + \frac{\nu}{\xi} e^{-(\alpha - \frac{1}{2})(R - y)}\right),$$

where

$$\lambda = \log\left(1 + \frac{\pi^2 K}{4}\right),$$

with K the constant from Lemma 1.6.2.

Proof. We will proceed in a similar fashion as for Lemma 4.6.4. That is, we will bound the expected number of common neighbours by the number of neighbors of p whose y-coordinate is above the intersection of the right boundary of $\mathcal{B}(p)$ and the left boundary of $\mathcal{B}(p')$. Denote by $\hat{y}$ the height of this intersection point. Then

$$\mathbb{E}\left[\mathcal{N}_{\mathrm{Po}}(p,p')\right] \leq \frac{2\alpha\nu}{\pi}\int_{\hat{y}}^{R-y}\Phi(y,y_1)e^{-\alpha y_1}\,\mathrm{d}y_1 + \alpha\nu e^{R/2}\int_{R-y}^{R} e^{-\alpha y_1}\,\mathrm{d}y_1.$$

The second integral is bounded by $\frac{\nu}{\xi}\mu\left(\mathcal{B}_{\mathrm{box}}(y)\right)e^{-(\alpha-\frac{1}{2})(R-y)}$. We bound the first integral using Lemma 1.6.2 as

$$\begin{aligned}\frac{2\alpha\nu}{\pi}\int_{\hat{y}}^{R-y}\Phi(y,y_1)e^{-\alpha y_1}\,\mathrm{d}y_1 &\leq \frac{2\alpha\nu}{\pi}\int_{\hat{y}}^{R-y}\left(e^{\frac{y+y_1}{2}} + Ke^{\frac{3}{2}(y+y_1)-R}\right)e^{-\alpha y_1}\,\mathrm{d}y_1\\ &\leq \frac{2\alpha\nu}{\pi}(1+K)e^{\frac{y}{2}}\int_{\hat{y}}^{R-y} e^{-(\alpha-\frac{1}{2})y_1}\,\mathrm{d}y_1\\ &\leq (1+K)\mu\left(\mathcal{B}_{\mathrm{box}}(p)\right)e^{-(\alpha-\frac{1}{2})\hat{y}},\end{aligned}$$

where we have used that $\frac{3y_1}{2} \leq R - y + \frac{y_1}{2}$ for all $y_1 \leq R-y$ for the second line.

It remains to compute $\hat{y}$, for which we will establish the bound

$$\hat{y} \geq 2\log\left(\frac{x'}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right) - 2\lambda. \tag{4.75}$$

To show (4.75) we note that for any point $y_1 \geq \hat{y}$, the corresponding x-coordinate of the left boundary of $\mathcal{B}(p')$ must be to the left of that of the ball $\mathcal{B}(p)$, i.e. $x' - \Phi(y',y_1) \leq \Phi(y,y_1)$. Therefore it is enough to show that for all

$$y_1 \leq 2\log\left(\frac{x'}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right) - 2\lambda,$$

it holds that $\Phi(y,y_1) \leq x' - \Phi(y',y_1)$, with λ as defined in the statement of the lemma. Note that by assumption on $|x-x'|$ the above upper bound is non-negative. Using Lemma 1.6.2 it suffices to prove that for all such y_1,

$$e^{\frac{y+y_1}{2}} + Ke^{\frac{3}{2}(y+y_1)-R} \leq x' - e^{\frac{y'+y_1}{2}} - Ke^{\frac{3}{2}(y'+y_1)-R},$$

which is equivalent to

$$\left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)e^{\frac{y_1}{2}} + Ke^{-R}\left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)^3 e^{\frac{3y_1}{2}} \leq x'.$$

Plugging the upper bound for y_1 into the left-hand side and using that $(e^{y/2} + e^{y'/2})^3 \geq e^{3y/2} + e^{3y'/2}$, we obtain

$$\left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)e^{\frac{y_1}{2}} + Ke^{-R}\left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)^3 e^{\frac{3y_1}{2}} \leq x'e^{-\lambda} + Ke^{-R}(x')^3e^{-3\lambda}$$

$$\leq x'\left(e^{-\lambda} + \frac{\pi^2 K}{4} e^{-3\lambda}\right)$$
$$\leq x' e^{-\lambda}\left(1 + \frac{\pi^2 K}{4}\right) = x',$$

where we have also used that $x' \leq \frac{\pi}{2} e^{-R/2}$. □

Degrees

We now return to the joint degree distribution of vertices in G_{box}. Recall the definition of $\mathcal{E}_\varepsilon(k_n)$ as

$$\mathcal{E}_\varepsilon(k_n) = \left\{(p, p') \in \mathcal{R} \times \mathcal{R} \,:\, y, y' \in \mathcal{K}_C(k_n) \text{ and } |x - x'|_n > k_n^{1+\varepsilon}\right\}.$$

The following result, which follows from Lemma 4.6.4, shows that on this set, the expected number of common neighbours is $o\,(k_n)$.

Lemma 4.6.6. *Fix $0 < \varepsilon < 1$ and let $\varepsilon' = \min\{\varepsilon(2\alpha - 1), \varepsilon\}$. Then for all $(p, p') \in \mathcal{E}_\varepsilon(k_n)$, as $n \to \infty$,*

$$\mu\left(\mathcal{B}_{box}\left(p\right) \cap \mathcal{B}_{box}\left(p'\right)\right) = O\left(k_n^{1-\varepsilon'}\right).$$

Proof. Since $\mu\left(\mathcal{B}_{\mathrm{box}}\left(p\right)\right), \mu\left(\mathcal{B}_{\mathrm{box}}\left(p'\right)\right) = \Theta\left(k_n\right)$ for all $(p, p') \in \mathcal{E}_\varepsilon(k_n)$, Lemma 4.6.4 implies that

$$\mu\left(\mathcal{B}_{\mathrm{box}}\left(p\right) \cap \mathcal{B}_{\mathrm{box}}\left(p'\right)\right) \leq O\left(k_n\right)\phi_n(p, p'),$$

where

$$\phi_n(p, p') = 2\left(\frac{|x - x'|}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right)^{-(2\alpha-1)} + \frac{3\nu^{2\alpha+1} e^{-(\alpha-\frac{1}{2})R} e^{\alpha y}}{2\pi^{2\alpha}\left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)} + \frac{\nu e^{-(\alpha-\frac{1}{2})R}}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}. \tag{4.76}$$

We thus need to show that $\phi_n(p, p') = O\left(k_n^{-\varepsilon}\right)$. For $(p, p') \in \mathcal{E}_\varepsilon(k_n)$, it holds that $e^{y/2}, e^{y'/2} = \Theta\left(k_n\right)$ and $|x - x'| > k_n^{1+\varepsilon}$ and hence

$$2\left(\frac{|x - x'|}{e^{\frac{y}{2}} + e^{\frac{y'}{2}}}\right)^{-(2\alpha-1)} = O\left(k_n^{-\varepsilon(2\alpha-1)}\right).$$

For the second term in $\phi_n(p, p')$ we use that $e^{\alpha y} = \Theta\left(k_n^{2\alpha}\right)$ and $e^R = \Theta\left(n^2\right)$ to obtain

$$\frac{3\nu^{2\alpha+1} e^{-(\alpha-\frac{1}{2})R} e^{\alpha y}}{2\pi^{2\alpha}\left(e^{\frac{y}{2}} + e^{\frac{y'}{2}}\right)} = O\left(1\right) n^{-(2\alpha-1)} k_n^{2\alpha-1} = O\left(n^{-(\alpha-\frac{1}{2})}\right).$$

Finally, the third term in (4.76) is $O\left(n^{-(2\alpha-1)} k_n^{-1}\right)$, and we conclude that

$$\phi_n(p, p') = O\left(k_n^{-\varepsilon(2\alpha-1)} + n^{-(\alpha-\frac{1}{2})} + n^{-(2\alpha-1)} k_n^{-1}\right) = O\left(k_n^{-\varepsilon'}\right),$$

where we have used that $\varepsilon' = \min\{\varepsilon(2\alpha - 1), \varepsilon\}$. □

It is clear that using Lemma 4.6.5 instead of Lemma 4.6.4, the above proof applies to the Poissonized KPKVB model, yielding the following result:

Lemma 4.6.7. *Fix $0 < \varepsilon < 1$ and let $\varepsilon' = \min\{\varepsilon(2\alpha - 1), \varepsilon\}$. Then for all $(p, p') \in \mathcal{E}_\varepsilon(k_n)$, as $n \to \infty$,*

$$\mu\left(\mathcal{B}\left(p\right) \cap \mathcal{B}\left(p'\right)\right) = O\left(k_n^{1-\varepsilon'}\right).$$

As a corollary we get that on the set $\mathcal{E}_\varepsilon(k_n)$ the joint degree distribution in G_{box} is asymptotically equivalent to the product of the degree distributions. We shall however prove a slightly stronger result (Lemma 4.6.9) which also takes care of bounded shifts in the joint degree distribution $\rho_{\text{box}}(p, p', k_n - t, k_n - t')$, for some uniformly bounded $t, t' \in \mathbb{Z}$. For this we first need the following simple result for Poisson distributions:

Lemma 4.6.8. *Let $k_n \to \infty$ be a sequence of non-negative integers and $X = \text{Po}(\lambda_n)$ be a Poisson random variable with mean λ_n satisfying*

$$k_n - C\sqrt{k_n \log(k_n)} \leq \lambda_n \leq k_n + C\sqrt{k_n \log(k_n)}$$

for some $C > 0$. Then, for any $t_n, s_n = O(1)$, as $n \to \infty$,

$$\mathbb{P}\left(X = k_n - t_n\right) \sim \mathbb{P}\left(X = k_n - s_n\right).$$

Proof. Note that $k_n > t_n, s_n$ for large enough n. Hence, using Stirling's formula, as $n \to \infty$,

$$\begin{aligned}
\frac{\mathbb{P}\left(X = k_n - t_n\right)}{\mathbb{P}\left(X = k_n - s_n\right)} &= \frac{(k_n - t_n - (s_n - t_n))!}{(k_n - t_n)!}\lambda_n^{s_n - t_n} \\
&\sim \sqrt{\frac{k_n - s_n}{k_n - t_n}}\,\frac{(k_n - s_n)^{k_n - s_n}}{(k_n - t_n)^{k_n - t_n}}\,e^{t_n - s_n}\lambda_n^{s_n - t_n} \\
&= \sqrt{\ell_n}(\ell_n)^{k_n - t_n}e^{t_n - s_n}(k_n - s_n)^{t_n - s_n}\lambda_n^{s_n - t_n} \\
&= \sqrt{\ell_n}e^{(k_n - t_n)\log(\ell_n) + t_n - s_n}\left(\frac{k_n - s_n}{\lambda_n}\right)^{t_n - s_n},
\end{aligned}$$

where we wrote $\ell_n = (k_n - s_n)/(k_n - t_n)$. Note that $\ell_n \to 1$ and hence $\sqrt{\ell_n} \to 1$. Moreover, since $(k_n - s_n)/\lambda_n \to 1$ and $|s_n - t_n| = O(1)$, we have that $\left(\frac{k_n - s_n}{\lambda_n}\right)^{t_n - s_n} \sim 1$. Therefore it remains to show that

$$\lim_{n\to\infty} e^{(k_n - t_n)\log(\ell_n) + t_n - s_n} = 1.$$

For this we note that for any x, such that $|x| \leq 1/2$, we have

$$x - x^2 \leq \log(1 + x) \leq x.$$

Write $x_n = \ell_n - 1 = \frac{t_n - s_n}{k_n - t_n}$. Then by the assumptions of the lemma, $x_n \to 0$, and thus, for n large enough,

$$t_n - s_n - \frac{(t_n - s_n)^2}{k_n - t_n} \leq (k_n - t)\log(\ell_n) \leq t_n - s_n.$$

In particular

$$e^{-\frac{(t_n - s_n)^2}{k_n - t_n}} \leq e^{(k_n - t_n)\log(\ell_n) + t_n - s_n} \leq 1,$$

and the result follows since $\frac{(t_n - s_n)^2}{k_n - t_n} \to 0$. $\square$

We can now prove the main result of this section:

Lemma 4.6.9. *Let $0 < \varepsilon < 1$, $k_n \to \infty$ and let $t_n, t'_n, s_n, s'_n \in \mathbb{Z}$ be uniformly bounded. Then for any $(p, p') \in \mathcal{E}_\varepsilon(k_n)$, as $n \to \infty$,*

$$\rho_{\text{box}}(p, p', k_n - t_n, k_n - t'_n) = (1 + o(1))\rho_{\text{box}}(p, k_n - s_n)\rho_{\text{box}}(p', k_n - s'_n).$$

Proof. Define the random variables

$$\begin{aligned} X_1(p,p') &:= \text{Po}\left(\mu\left(\mathcal{B}_{\text{box}}(p) \setminus \mathcal{B}_{\text{box}}(p')\right)\right), \\ X_2(p,p') &:= \text{Po}\left(\mu\left(\mathcal{B}_{\text{box}}(p') \setminus \mathcal{B}_{\text{box}}(p)\right)\right), \\ Y(p,p') &:= \text{Po}\left(\mu\left(\mathcal{B}_{\text{box}}(p) \cup \mathcal{B}_{\text{box}}(p')\right)\right), \end{aligned}$$

so that

$$\begin{aligned} &\rho_{\text{box}}(p, p', k_n - t_n, k_n - t'_n) \\ &= \mathbb{P}\left(X_1(p,p') + Y(p,p') = k_n - t_n, X_2(p,p') + Y(p,p') = k_n - t'_n\right). \end{aligned}$$

Since by Lemma 4.6.6 it holds that $\mu\left(\mathcal{B}_{\text{box}}(p) \cap \mathcal{B}_{\text{box}}(p')\right) = O\left(k_n^{1-\varepsilon'}\right)$, it follows from Lemma 4.6.3 that

$$\rho_{\text{box}}(p, p', k_n - t_n, k_n - t'_n) = (1 + o(1))\rho_{\text{box}}(p, k_n - t_n)\rho_{\text{box}}(p', k_n - t'_n).$$

The result then follows by applying Lemma 4.6.8 twice. $\square$

4.6.3 Concentration result for main triangle contribution

We now turn to Proposition 4.6.1. Before we dive into the proof let us first give a high level overview of the strategy and the flow of the arguments.

Recall (see (4.62)) that for any $C > 0$

$$\widetilde{T}_{\text{box}}(k_n, C) = \sum_{p \in \mathcal{P}_n \cap \mathcal{K}_{C,n}(k_n)} \mathbb{1}_{\{\deg_{\text{box}}(p) = k\}} \widetilde{T}_{\text{box}}(p).$$

Then we have

$$\widetilde{T}_{\text{box}}(k_n, C)^2 = \sum_{p,p' \in \mathcal{P}_n \cap \mathcal{K}_C(k_n)} \mathbb{1}_{\{\deg_{\text{box}}(p),\, \deg_{\text{box}}(p')=k_n\}} \times \sum_{(p_1,p_2),(p_1',p_2') \in \mathcal{P}_n}^{\neq} \widetilde{T}_{\mathcal{P}}(p, p_1, p_2)\widetilde{T}_{\mathcal{P}}(p', p_1', p_2').$$

This expression can be written as the sum of several terms, depending on how $\{p, p_1, p_2\}$ and $\{p', p_1', p_2'\}$ intersect. To this end we define, for $a \in \{0, 1\}$ and $b \in \{0, 1, 2\}$,

$$I_{a,b} = \sum_{\substack{p,p' \in \mathcal{P}_n \cap \mathcal{K}_C(k) \\ |\{p\} \cap \{p'\}| = a}} \mathbb{1}_{\{\deg_{\text{box}}(p),\, \deg_{\text{box}}(p')=k_n\}} J_b(p, p'),$$

where

$$J_b(p, p') = \sum_{\substack{p_1,p_2,p_1',p_2' \in \mathcal{P}_n \\ |\{p_1,p_2\} \cap \{p_1',p_2'\}| = b}}^{\neq} T_{\mathcal{P},n}(p, p_1, p_2) T_{\mathcal{P},n}(p', p_1', p_2'),$$

with the sum taken over all two distinct pairs (p_1, p_2) and (p_1', p_2'). Then we have

$$\widetilde{T}_{\text{box}}(k, C)^2 = \sum_{a=0}^{1} \sum_{b=0}^{2} I_{a,b}.$$

To prove Proposition 4.6.1 we will deal with each of the $I_{a,b}$ separately, showing that

$$\mathbb{E}[I_{0,0}] = (1 + o(1))\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]^2, \tag{4.77}$$

and for all other combinations

$$\mathbb{E}[I_{a,b}] = o\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]^2\right). \tag{4.78}$$

Note that $I_{1,2} = \widetilde{T}_{\text{box}}(k_n, C)$ and since (4.64) implies that $\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right] \to \infty$, it follows that (4.78) holds for $I_{1,2}$.

Recall that $\mathcal{R}(k_n, C) = [-I_n, I_n] \times \mathcal{K}_C(k_n)$ and (4.66)

$$\mathcal{E}_\varepsilon(k_n) = \left\{(p, p') \in \mathcal{R} \times \mathcal{R} \,:\, y, y' \in \mathcal{K}_C(k_n) \text{ and } |x - x'|_n > k_n^{1+\varepsilon}\right\}.$$

Let $\mathcal{E}_\varepsilon(k_n)^c$ be the same set but with $|x - x'|_n \le k_n^{1+\varepsilon}$ and denote by $I^*_{a,b}$ the part of $I_{a,b}$ where $(p, p') \in \mathcal{E}_\varepsilon(k_n)$. We split the analysis between $I^*_{a,b}$ and $I_{a,b} - I^*_{a,b}$. The idea for these two cases is that by Lemma 4.6.9 it follows that on the set $\mathcal{E}_\varepsilon(k_n)$ and for any uniformly bounded $t, t' \in \mathbb{Z}$, the joint degree distribution factorizes, i.e.

$$\rho_{\text{box}}(p, p', k_n + t, k_n + t') = (1 + o(1))\rho_{\text{box}}(p, k_n)\rho_{\text{box}}(p, k_n).$$

In particular this allows us to prove that $\mathbb{E}\left[I^*_{0,0}\right] = (1+o(1))\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]^2$.

On the other hand, the expected number of points in $\mathcal{E}_\varepsilon(k_n)^c$ is indeed in $O\left(k_n^{1+\varepsilon}k_n^{-2\alpha}\mathbb{E}\left[N_{\text{box}}(k_n)\right]\right) = o\left(\mathbb{E}\left[N_{\text{box}}(k_n)\right]^2\right)$, where the latter is the expected number of points in $\mathcal{R}(k_n, C) \times \mathcal{R}(k_n, C)$. Hence we expect the contributions coming from $\mathcal{E}_\varepsilon(k_n)^c$ to be negligible.

Proof of Proposition 4.6.1. Throughout this proof we set $i = |\{p', p_1, p_2, p_1', p_2'\} \cap \mathcal{B}_{\text{box}}(p)|$, $j = |\{p'\} \cap \mathcal{B}_{\text{box}}(p)|$ and define i', j' in a similar way by interchanging the primed and non-primed variables. In addition, we write $\widetilde{D}_{\text{box}}(p, p', k, \ell)$ to denote the indicator that $|\mathcal{B}_{\text{box}}(p) \cap (\mathcal{P} \setminus \{p, p', p_1, p_2, p_1', p_2'\})| = k$ and $|\mathcal{B}_{\text{box}}(p') \cap (\mathcal{P} \setminus \{p, p', p_1, p_2, p_1', p_2'\})| = \ell$. Note that this also depends on $\{p_1, p_2, p_1', p_2'\}$ but we suppressed this to keep notation concise. Similarly we write $D_{\text{box}}(p, p', k, \ell)$ to denote the indicator that $|\mathcal{B}_{\text{box}}(p) \cap (\mathcal{P} \setminus \{p, p'\})| = k$ and $|\mathcal{B}_{\text{box}}(p') \cap (\mathcal{P} \setminus \{p, p'\})| = \ell$, which now only depends on p and p'. Then, by the Campbell-Mecke formula,

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{\deg_{\text{box}}(p) = k_n, \deg_{\text{box}}(p') = k_n\}} J_b(p, p')\right] \\
&= \mathbb{E}\left[\sum_{\substack{p_1, p_2, p_1', p_2' \in \mathcal{P}_n \\ |\{p_1, p_2\} \cap \{p_1', p_2'\}| = b}}^{\neq} \widetilde{D}_{\text{box}}(p, p', k_n - i, k_n - i')\, \widetilde{T}_{\text{box}}(p, p_1, p_2) \widetilde{T}_{\text{box}}(p', p_1', p_2')\right],
\end{aligned}
$$

where the sum is over all distinct pairs (p_1, p_2) and (p_1', p_2'). We also know that

$$
\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n)\right] = \Theta\left(n k_n^{-(2\alpha - 1)} s(k_n)\right).
$$

We will now proceed to establish (4.77) and (4.78).

Computing $I_{0,0}$ We first show that

$$
\mathbb{E}\left[I_{0,0} - I^*_{0,0}\right] = o\left(\mathbb{E}\left[T_{\text{box}}(k_n, C)\right]^2\right), \tag{4.79}
$$

so that for the remainder of the proof we only need to consider $p, p' \in \mathcal{E}_\varepsilon(k_n)$ and hence, we can apply Lemma 4.6.9.

For J_0 we have, using Lemma 4.6.9,

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{\deg_{\text{box}}(p) = k_n, \deg_{\text{box}}(p') = k_n\}} J_0(p, p')\right] \\
&= \mathbb{E}\left[\sum_{\substack{p_1, p_2, p_1', p_2' \in \mathcal{P} \setminus \{p, p'\} \\ |\{p_1, p_2\} \cap \{p_1', p_2'\}| = 0}}^{\neq} \widetilde{D}_{\text{box}}(p, p', k_n - i, k_n - i')\, \widetilde{T}_{\text{box}}(p, p_1, p_2) \widetilde{T}_{\text{box}}(p', p_1', p_2')\right] \\
&= \mathbb{E}\left[D_{\text{box}}(p, p', k_n - j - 2, k_n - j' - 2) \sum_{p_1, p_2 \in \mathcal{P} \setminus p}^{\neq} \widetilde{T}_{\text{box}}(p, p_1, p_2)\right.
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times \sum_{p_1',p_2' \in \mathcal{P}\setminus p'}^{\neq} \widetilde{T}_{\text{box}}(p',p_1',p_2')\Bigg] \\
&= (1+o(1))\rho_{\text{box}}(p,p',k_n,k_n)\mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\middle|\deg_{\text{box}}(p)=k_n\right] \\
&\qquad\qquad\qquad\qquad\qquad \times \mathbb{E}\left[\widetilde{T}_{\text{box}}(p')\middle|\deg_{\text{box}}(p')=k_n\right].
\end{aligned}
$$

Next we recall that for all $y' \in \mathcal{K}_C(k_n)$ (see (4.41)),

$$
\mathbb{E}\left[\widetilde{T}_{\text{box}}(p')\middle|\deg_{\text{box}}(p')=k_n\right] = \binom{k_n}{2}\mu\left(\mathcal{B}_{\text{box}}\left(p'\right)\right)^{-2}\mathbb{E}\left[\widetilde{T}_{\text{box}}(p')\right] = O\left(1\right)k_n^2 P(y'),
$$

where $p' = (x',y')$ and we have used that $\mathbb{E}\left[\widetilde{T}_{\text{box}}(p')\right] = (1+o(1))k_n^2 P(y')$, for all $y' \in \mathcal{K}_C(k_n)$. Therefore, using that $\rho_{\text{box}}(p,p',k_n,k_n) \le \rho_{\text{box}}(p,k_n)$,

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{\deg_{\text{box}}(p)=k_n,\deg_{\text{box}}(p')=k_n\}}J_0(p,p')\right] \\
&\le O\left(k_n^2\right)\rho_{\text{box}}(p,k_n)\mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\middle|\deg_{\text{box}}(p)=k_n\right]P(y'),
\end{aligned}
$$

and thus

$$
\begin{aligned}
&\mathbb{E}\left[I_{0,0}-I_{0,0}^*\right] \\
&= \int_{\mathcal{E}_\varepsilon(k_n)^c}\mathbb{E}\left[\mathbb{1}_{\{\deg_{\text{box}}(p),\deg_{\text{box}}(p')=k_n\}}J_0(p,p')\right]f(x,y)f(x',y')\,dx'\,dx\,dy'\,dy \\
&\le O\left(k_n^2\right)k_n^{1+\varepsilon}\left(\int_{a_n^-}^{a_n^+}P(y')e^{-\alpha y'}\,dy'\right)\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right] \\
&= O\left(k_n^{3+\varepsilon-2\alpha}s_\alpha(k_n)\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right]\right) \\
&= o\left(nk_n^{-(2\alpha-1)}s_\alpha(k_n)\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right]\right) = o\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right]^2\right),
\end{aligned}
$$

which proves (4.79). For the last line we have used that $k_n^{2+\varepsilon} = o\left(n\right)$ and

$$
\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right] = \Theta\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n)\right]\right) = \Theta\left(nk_n^{-(2\alpha-1)}s_\alpha(k_n)\right).
$$

We will now show that

$$
\mathbb{E}\left[I_{0,0}^*\right] = (1+o\left(1\right))\mathbb{E}\left[\mathbb{E}\left[T_{\text{box}}(k_n,C)\right]^2\right].
$$

Recall the result from Lemma 4.6.9, that for $(p,p') \in \mathcal{E}_\varepsilon(k_n)$ and any two uniformly bounded $t,t' \in \mathbb{Z}$,

$$
\rho_{\text{box}}(p,p',k_n+t,k_n+t') = (1+o\left(1\right))\rho_{\text{box}}(p,k_n)\rho_{\text{box}}(p,k_n).
$$

Therefore, by defining $h(y) = \mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(y)\middle|\, \mathrm{deg}_{\mathrm{box}}(y) = k_n\right]$

$$
\begin{aligned}
&\mathbb{E}\left[I^*_{0,0}\right] \\
&= (1 + o(1)) \int_{\mathcal{E}_\varepsilon(k_n)} \rho_{\mathrm{box}}(p, k_n)\rho_{\mathrm{box}}(p', k_n)h(y)h(y')f(x, y)f(x', y')\, dx'\, dx\, dy'\, dy.
\end{aligned}
$$

The difference with $\mathbb{E}\left[\mathbb{E}\left[T_{\mathrm{box}}(k_n, C)\right]^2\right]$ is that the above integral is over $\mathcal{E}_\varepsilon(k_n)$ instead of $\mathcal{R}(k_n, C)\times\mathcal{R}(k_n, C)$. Since the difference between the two sets is $\mathcal{E}_\varepsilon(k_n)^c$ and $nk_n^{1+\varepsilon} = o\left(n^2\right)$ it follows that

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[T_{\mathrm{box}}(k_n, C)\right]^2\right] \\
&\qquad - \int_{\mathcal{E}_\varepsilon(k_n)} \rho_{\mathrm{box}}(p, k_n)\rho_{\mathrm{box}}(p', k_n)h(y)h(y')f(x, y)f(x', y')\, \mathrm{d}x'\, \mathrm{d}x\, \mathrm{d}y'\, \mathrm{d}y \\
&= \int_{\mathcal{E}_\varepsilon(k_n)^c} \rho_{\mathrm{box}}(p, k_n)\rho_{\mathrm{box}}(p', k_n)h(y)h(y')f(x, y)f(x', y')\, \mathrm{d}x'\, \mathrm{d}x\, \mathrm{d}y'\, \mathrm{d}y \\
&= O\left(k_n^{1+\varepsilon} n\right)\left(\int_{\mathcal{K}_C(k_n)} h(y)\rho_{\mathrm{box}}(y, k_n)\alpha e^{-\alpha y}\, \mathrm{d}y\right)^2 = o\left(\mathbb{E}\left[\mathbb{E}\left[T_{\mathrm{box}}(k_n, C)\right]^2\right]\right).
\end{aligned}
$$

Thus we conclude that $\mathbb{E}\left[I^*_{0,0}\right] = (1 + o(1))\mathbb{E}\left[\mathbb{E}\left[T_{\mathrm{box}}(k_n, C)\right]^2\right]$, which finishes the proof of (4.77).

Computing $\mathbb{E}\left[I_{0,1}\right]$ We first write

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{\mathrm{deg}_{\mathrm{box}}(p)=k_n, \mathrm{deg}_{\mathrm{box}}(p')=k_n\}} J_1\right] \\
&\leq O(1)\, k_n \rho_{\mathrm{box}}(p, p', k_n, k_n)\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(p)\middle|\, \mathrm{deg}_{\mathrm{box}}(p) = k_n\right]
\end{aligned}
\tag{4.80}
$$

Then, using that $\rho_{\mathrm{box}}(p, p', k_n, k_n) \leq \rho_{\mathrm{box}}(p, k_n)$,

$$
\begin{aligned}
&\mathbb{E}\left[I_{0,1} - I^*_{0,1}\right] \\
&\int_{\mathcal{E}_\varepsilon(k_n)^c} \mathbb{E}\left[\mathbb{1}_{\{\mathrm{deg}_{\mathrm{box}}(p), \mathrm{deg}_{\mathrm{box}}(p')=k_n\}} J_1(p, p')\right] f(x, y)f(x', y')\, dx'\, dx\, dy'\, dy \\
&= k_n \int_{\mathcal{K}_C(k_n)^2} \mathbb{1}_{\left\{|x-x'|\leq k_n^{1+\varepsilon}\right\}}\rho_{\mathrm{box}}(p, k_n) \\
&\qquad\qquad \times \mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(p)\middle|\, \mathrm{deg}_{\mathrm{box}}(p) = k_n\right] f(x, y)f(x', y')\, dx'\, dx\, dy'\, dy \\
&\leq O\left(k_n^{2+\varepsilon}\right)\left(\int_{a_n^-}^{a_n^+} e^{-\alpha y'}\, dy'\right)\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(k_n, C)\right] \\
&= O\left(k_n^{2+\varepsilon-2\alpha}\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(k_n, C)\right]\right).
\end{aligned}
$$

Recall that $\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(k_n, C)\right] = \Theta\left(nk_n^{-(2\alpha-1)}s(k_n)\right)$. Therefore in order to show that $\mathbb{E}\left[I_{0,1} - I^*_{0,1}\right] = o\left(\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(k_n, C)\right]^2\right)$, it suffices to show that $k_n^{2+\varepsilon-2\alpha} = o\left(nk_n^{-(2\alpha-1)}s(k_n)\right)$. When $\frac{1}{2} < \alpha \leq \frac{3}{4}$ we have

$$\frac{4\alpha - 1 + \varepsilon}{2\alpha + 1} < 1,$$

for ε small enough. Hence

$$n^{-1}k_n^{2\alpha-1}s(k_n)^{-1}k_n^{2+\varepsilon-2\alpha} = n^{-1}k_n^{4\alpha-1+\varepsilon} = o\left(n^{-1}n^{\frac{4\alpha-1+\varepsilon}{2\alpha+1}}\right) = o\left(1\right).$$

When $\alpha \geq \frac{3}{4}$,

$$n^{-1}k_n^{2\alpha-1}s(k_n)^{-1}k_n^{2+\varepsilon-2\alpha} = O\left(\log(k_n)\right)n^{-1}k_n^{2+\varepsilon} = o\left(1\right),$$

for ε small enough.

For $(p, p') \in \mathcal{E}_\varepsilon(k_n)$ we assume without loss of generality that $p'_1 = p_1 = (x_1, y_1)$, i.e.

$$J_{0,1} = \sum_{(p_1,p_2)\in\mathcal{P}\setminus\{p\}}^{\neq} \widetilde{T}_{\mathrm{box}}(p, p_1, p_2) \sum_{p'_2\in\mathcal{P}\setminus\{p',p_1\}} \widetilde{T}_{\mathrm{box}}(p', p_1, p'_2).$$

Now let $Z_{0,1}$ denote the part of $J_{0,1}$ where $y_1 \leq 4\log(k_n)$ and $y_2, y'_2 \leq \varepsilon\log(k_n)$.

We first analyze $\mathbb{E}\left[Z_{0,1}|\deg_{\mathrm{box}}(p), \deg_{\mathrm{box}}(p') = k_n\right]$. When $y_1 \leq 4\log(k_n)$ and both $y_2, y'_2 \leq \varepsilon\log(k_n)$ we have that

$$|x_2 - x'_2| \leq |x_1 - x_2| + |x_1 - x'_2| \leq e^{\frac{y_1}{2}}\left(e^{\frac{y_2}{2}} + e^{\frac{y'_2}{2}}\right) \leq 2k_n^{2+\varepsilon},$$

whenever $\widetilde{T}_{\mathrm{box}}(p, p_1, p_2)\widetilde{T}_{\mathrm{box}}(p', p_1, p'_2) > 0$ while both $|x - x_2|, |x' - x'_2| = O\left(k_n^{1+\varepsilon}\right)$. Hence it follows that $\widetilde{T}_{\mathrm{box}}(p, p_1, p_2)\widetilde{T}_{\mathrm{box}}(p', p_1, p'_2) > 0$ implies that

$$|x - x'| \leq |x - x_2| + |x_2 - x'_2| + |x'_2 - x'| = O\left(k_n^{2+\varepsilon}\right).$$

Next, by integrating only over x'_2 and y'_2, we get

$$\begin{aligned}
&\mathbb{E}\left[Z_{0,1}|\deg_{\mathrm{box}}(p), \deg_{\mathrm{box}}(p') = k_n\right] \\
&= O\left(e^{\frac{y'}{2}}\mathbb{1}_{\left\{|x-x'|\leq O(1)k_n^{2+\varepsilon}\right\}}\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(p)\middle|\deg_{\mathrm{box}}(p) = k_n\right]\right) \\
&= O\left(k_n\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(p)\middle|\deg_{\mathrm{box}}(p) = k_n\right]\right).
\end{aligned}$$

Thus

$$\int_{\mathcal{E}_\varepsilon(k_n)} \rho_{\mathrm{box}}(p, p', k_n, k_n)$$

$$
\begin{aligned}
&\quad \times \mathbb{E}\left[Z_{0,1} \middle| \deg_{\text{box}}(p), \deg_{\text{box}}(p') = k_n \right] f(x,y) f(x',y') \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}x' \, \mathrm{d}y' \\
&= O\left(k_n^{3+\varepsilon}\right) \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right] \int_{\mathcal{K}_C(k_n)} \rho_{\text{box}}(y', k_n) e^{-\alpha y'} \, \mathrm{d}y' \\
&= O\left(k_n^{2+\varepsilon} k_n^{-2\alpha} \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]\right) = o\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]^2\right),
\end{aligned}
$$

where the last line follows from the analysis done for $\mathbb{E}\left[I_{0,0} - I_{0,0}^*\right]$.

It now remains to consider $J_{0,1} - Z_{0,1} := Z_{0,1}^*$. We will show that

$$
\mathbb{E}\left[Z_{0,1}^* \middle| \deg_{\text{box}}(p), \deg_{\text{box}}(p') = k_n \right] = o\left(k_n^4 s(k_n)^2\right). \tag{4.81}
$$

Using that the joint degree distribution factorizes on $\mathcal{E}_\varepsilon(k_n)$ this then implies that

$$
\begin{aligned}
\mathbb{E}\left[I_{0,1}^*\right] &= o\left(k_n^4 s(k_n)^2\right) \left(\int_{\mathcal{R}(k_n, C)} \rho_{\text{box}}(y, k_n) f(x,y) \, \mathrm{d}x \, \mathrm{d}y \right)^2 \\
&= o\left(\left(n s(k_n) k_n^{-2\alpha+1}\right)^2\right) = o\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n, C)\right]^2\right),
\end{aligned}
$$

which finishes the proof of (4.78) for $a = 0, b = 1$.

We first consider the part with $y_1 > 4\log(k_n)$. Since the integration over x_1, x_2 and x_2' of $\mathbb{E}\left[Z_{0,1}^* \middle| \deg_{\text{box}}(p), \deg_{\text{box}}(p') = k_n \right]$ is bounded by $O\left(e^y e^{\frac{y'}{2}}\right)$ we get that the contribution to $\mathbb{E}\left[Z_{0,1}^* \middle| \deg_{\text{box}}(p), \deg_{\text{box}}(p') = k_n \right]$ due to $y > 4\log(k_n)$ and $(p, p') \in \mathcal{E}_\varepsilon(k_n)$ is

$$
\begin{aligned}
O\left(e^y e^{\frac{y'}{2}} \int_{4\log(k_n)}^R e^{-(\alpha - \frac{1}{2}) y_1} \, dy_1 \right) &= O\left(k_n^3 \int_{4\log(k_n)}^R e^{-(\alpha - \frac{1}{2}) y_1} \, dy_1 \right) \\
&= O\left(k_n^{3-(4\alpha-2)}\right) = o\left(k_n^4 s_\alpha(k_n)^2\right).
\end{aligned}
$$

Here the last step follows since, for $\frac{1}{2} < \alpha < \frac{3}{4}$,

$$
k_n^{3-(4\alpha-2)-4} s(k_n)^{-2} = k_n^{3-(4\alpha-2)-4+2(4\alpha-2)} = k_n^{-5+4\alpha} = o(1),
$$

while, for $\alpha = \frac{3}{4}$,

$$
k_n^{3-(4\alpha-2)-4} s(k_n)^{-2} = O\left(\log(k_n)^{-2}\right) k_n^{3-(4\alpha-2)-2} = O\left(\log(k_n)^{-2}\right) = o(1),
$$

and, for $\alpha > \frac{3}{4}$,

$$
k_n^{3-(4\alpha-2)-4} s(k_n)^{-2} = k_n^{3-(4\alpha-2)-2} = o(1).
$$

Next we consider the case where $y_1 \leq 4\log(k_n)$ and at least one of y_2, y_2' is larger than $\varepsilon \log(k_n)$. Due to symmetry it is enough to consider the case with $y_2 > \varepsilon \log(k_n)$. Here the contribution to $\mathbb{E}\left[Z_{0,1}^* \middle| \deg_{\text{box}}(p), \deg_{\text{box}}(p') = k_n \right]$ is

$$
\mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\right] O\left(e^{\frac{y'}{2}} \int_{\varepsilon \log(k_n)}^R e^{-(\alpha - \frac{1}{2}) y_2} \, dy_2 \right) = O\left(k_n^{1-\varepsilon(\alpha - \frac{1}{2})}\right) \mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\right]
$$

$$= O\left(k_n^{3-\varepsilon(\alpha-\frac{1}{2})} s(k_n)\right) = o\left(k_n^4 s(k_n)^2\right).$$

The last line follows since $k_n^{-1} = o(s(k_n))$ for $\frac{1}{2} < \alpha < \frac{3}{4}$ and $k_n^{-1} = O(s(k_n))$ for $\alpha \geq \frac{3}{4}$.

Computing $\mathbb{E}[I_{0,2}]$ In this case we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\{\deg_{\text{box}}(p)=k_n, \deg_{\text{box}}(p')=k_n\}} J_2\right] \\
&\quad = (1+o(1))\rho_{\text{box}}(p,p',k_n,k_n)\mathbb{E}\left[\widetilde{T}_{\text{box}}(p)\middle| \deg_{\text{box}}(p) = k_n\right].
\end{aligned}$$

We then use that $\rho_{\text{box}}(p,p',k_n,k_n) \leq \rho_{\text{box}}(p,k_n)$ to obtain

$$\begin{aligned}
\mathbb{E}\left[I_{0,2} - I_{0,2}^*\right] &= O\left(k_n^{1+\varepsilon}\right)\left(\int_{\mathcal{K}_C(k_n)} e^{-\alpha y'} \, \mathrm{d}y'\right) \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right] \\
&= O\left(k_n^{\varepsilon-(2\alpha-1)}\right) \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right] = o\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right]\right)
\end{aligned}$$

were the last line follows since $\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right] = \Theta\left(nk_n^{-(2\alpha-1)} s(k_n)\right)$ and $k_n^\varepsilon n^{-1} = o(s(k_n))$.

For the other term we use the fact that the degree distribution factorizes, i.e.

$$\begin{aligned}
\mathbb{E}\left[I_{0,2}^*\right] &= O(1)\left(\int_{\mathcal{R}(k_n,C)} \rho_{\text{box}}(y',k_n) f(x',y') \, \mathrm{d}x' \, \mathrm{d}y'\right) \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right] \\
&= O\left(nk_n^{-(2\alpha+1)}\right) \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right] = o\left(\mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right]^2\right),
\end{aligned}$$

where we have also used that $k_n^{-2} = o(s(k_n))$.

Computing $\mathbb{E}[I_{1,1}]$ Using (4.81) we get

$$\begin{aligned}
\mathbb{E}[I_{1,1}] &= O(k_n) \int_{\mathcal{R}(k_n,C)} \rho_{\text{box}}(y,k_n)\mathbb{E}\left[\widetilde{T}_{\text{box}}\middle| \deg_{\text{box}}(p) = k_n\right] f(x,y) \, \mathrm{d}x \, \mathrm{d}y \\
&= O(k_n) \, \mathbb{E}\left[\widetilde{T}_{\text{box}}(k_n,C)\right].
\end{aligned}$$

Now observe that, for $\frac{1}{2} < \alpha < \frac{3}{4}$,

$$k_n n^{-1} k_n^{(2\alpha-1)} s(k_n)^{-1} = k_n^{6\alpha-2} n^{-1} = O\left(n^{\frac{4\alpha-3}{2\alpha+1}}\right) = o(1),$$

while, for $\alpha \geq \frac{3}{4}$,

$$k_n n^{-1} k_n^{(2\alpha-1)} s(k_n)^{-1} = O\left(n^{-1} k_n^{-(2\alpha-1)}\right) = o(1).$$

We conclude that $k_n = o\left(nk^{-(2\alpha-1)}s(k_n)\right)$ and hence

$$\mathbb{E}\left[I_{1,1}\right] = o\left(\mathbb{E}\left[\widetilde{T}_{\mathrm{box}}(k_n, C)\right]^2\right).$$

□

4.7 Equivalence for local clustering in $G_{\mathbf{Po}}$ and G_{box}

In this section we establish the equivalence between $c^*(k; G_n)$ and $c^*(k; G_{\mathrm{box}})$ as expressed in Proposition 4.3.4, using the coupling procedure explained in Section 1.6.4. As in the previous section we write $|\cdot|_n$ for the norm $|\cdot|_{\pi e^{R/2}}$.

Recall the map Ψ from (1.7), i.e.,

$$\Psi(r, \theta) = \left(\theta \frac{e^{R/2}}{2}, R - r\right),$$

and that $\mathcal{B}(p)$ denotes the image under Ψ of the ball of hyperbolic radius R around the point $\Psi^{-1}(p)$. Under the coupling between the hyperbolic random graph and the finite box model, described in Section 1.6.4, two points $p = (x, y)$ and $p' = (x', y')$ are connected if and only if

$$|x - x'|_n \leq \Phi(y, y') = \frac{1}{2}e^{R/2} \arccos\left(\frac{\cosh(R-y)\cosh(R-y') - \cosh R}{\sinh(R-y)\sinh(R-y')}\right),$$

see (1.8). We will often use the result from Lemma 1.6.2 to approximate the function Φ, for $y + y' < R$, by

$$e^{\frac{1}{2}(y+y')} - Ke^{\frac{3}{2}(y+y')-R} \leq \Phi(R-y, R-y') \leq e^{\frac{1}{2}(y+y')} + Ke^{\frac{3}{2}(y+y')-R},$$

where K is a constant determined by the lemma.

4.7.1 Some results on the hyperbolic geometric graph

We start with some basic results for the hyperbolic random geometric graph. Recall that $\mathcal{B}_\infty(p) = \{p' \in \mathbb{R} \times \mathbb{R}_+ : |x - x'| \leq e^{(y+y')/2}\}$ and observe that (1.10) from Lemma 1.6.2 implies the following:

Corollary 4.7.1. *For sufficiently large n and $p \in \mathcal{R}$,*

$$\mathcal{B}_\infty(p) \cap \mathcal{R}([K, R]) \subseteq \mathcal{B}(p) \cap \mathcal{R}([K, R]),$$

where K is the constant from Lemma 1.6.2.

Furthermore, Lemma 1.6.2 enables us to determine the measure of a ball around a given point $p = (0, y)$ - this is will be fairly useful in our subsequent analysis.

Let $p \in \mathcal{R}$. Then we can see that the curve $x' = e^{\frac{1}{2}(y+y')}$ with $x' \geq 0$ meets the right boundary of $\mathcal{R}$, that is, the line $x' = \frac{\pi}{2}e^{R/2}$, at $y' = R - y + 2\ln\frac{\pi}{2}$. Hence, any point $p' \in \mathcal{R}([R-y+2\ln\frac{\pi}{2}, R])$ is included in $\mathcal{B}_\infty(p)$. In other words,

$$\mathcal{B}_\infty(p) \cap \mathcal{R}([R-y+2\ln\frac{\pi}{2}, R]) = \mathcal{R}([R-y+2\ln\frac{\pi}{2}, R]).$$

This, together with the fact that for any $u' = (r', \theta')$,

$$r' < y = R - r \Rightarrow d_{\mathbb{H}}(\Psi^{-1}(p), u') \leq R,$$

implies that

$$(\mathcal{B}(p) \triangle \mathcal{B}_\infty(p)) \cap \mathcal{R}([R-y+2\ln\frac{\pi}{2}, R]) = \emptyset, \tag{4.82}$$

where $A \triangle B$ denotes the symmetric difference of the sets A and B. We can now compute the expected number of points in $\mathcal{B}(p) \triangle \mathcal{B}_\infty(p)$, i.e. those that are a neighbour of p in only one of the two models.

Lemma 4.7.2. *Let $0 \leq y_n < R$ be such that $R - y_n \to \infty$ and write $p_n = (x_n, y_n)$. Then, as $n \to \infty$,*

$$\mu(\mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n)) = \Theta(1) \cdot \begin{cases} e^{(1/2-\alpha)R+\alpha y_n}, & \text{if } \alpha < 3/2, \\ (R-y_n)e^{3y/2-R}, & \text{if } \alpha = 3/2, \\ e^{3y_n/2-R}, & \text{if } \alpha > 3/2. \end{cases}$$

Proof. Let $r_n := R - y$. Lemma 1.6.2 implies that for such a p_n, if a point p belongs to $\mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n) \cap \mathcal{R}([0, r_n])$, then

$$|x_n - x| = \Theta(1) \cdot e^{\frac{3}{2}(y_n+y)-R}.$$

Now, if $p \in [r_n, r_n + 2\ln\frac{\pi}{2})]$ and also $p \in \mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n)$, then

$$|x_n - x|_n = \frac{\pi}{2}e^{R/2} - e^{\frac{1}{2}(y_n+y)}.$$

Finally, (4.82) implies that no point in $\mathcal{R}([r_n + 2\ln\frac{\pi}{2}, R])$ belongs to $\mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n)$. We first compute the expected number of points $p \in \mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n)$ that have $R - y \leq r_n$. The result depends on the value of α, yielding the following three cases:

$$\begin{aligned}\mu(\mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n) \cap \mathcal{R}([0, r_n])) &= \Theta(1) \cdot e^{3y_n/2-R} \int_0^{r_n} e^{(3/2-\alpha)y}\, dy \\ &= \Theta(1) \cdot \begin{cases} e^{(1/2-\alpha)R+\alpha y_n}, & \text{if } \alpha < 3/2, \\ (R-y_n)e^{3y_n/2-R}, & \text{if } \alpha = 3/2, \\ e^{3y_n/2-R}, & \text{if } \alpha > 3/2. \end{cases}\end{aligned}$$

Next we compute the number of remaining points in $\mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n)$ as

$$\begin{aligned}
\mu(\mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n) \cap \mathcal{R}([r_n, R])) &= \frac{\nu\alpha}{\pi} \int_{r_n}^{r_n + 2\ln\frac{\pi}{2}} \left(\frac{\pi}{2} e^{R/2} - e^{\frac{1}{2}(y_n + y)}\right) e^{-\alpha y}\, dy \\
&= O(1) \cdot e^{R/2} \int_{r_n}^{r_n + 2\ln\frac{\pi}{2}} e^{-\alpha y}\, dy \\
&= O(1) \cdot e^{R/2} e^{-\alpha r_n} = O(1) \cdot e^{(1/2-\alpha)R + \alpha y_n}.
\end{aligned}$$

Now note that for any $\alpha > 3/2$, we have

$$((1/2 - \alpha)R + \alpha y_n) - (3y_n/2 - R) = (3/2 - \alpha)(R - y_n) \to -\infty,$$

by our assumption on y_n. For $\alpha = 3/2$, these two quantities are equal. From these observations, we deduce that

$$\mu(\mathcal{B}(p_n) \triangle \mathcal{B}_\infty(p_n)) = \Theta(1) \cdot \begin{cases} e^{(1/2-\alpha)R + \alpha y_n}, & \text{if } \alpha < 3/2, \\ r_n e^{3y_n/2 - R}, & \text{if } \alpha = 3/2, \\ e^{3y_n/2 - R}, & \text{if } \alpha > 3/2. \end{cases}$$

□

4.7.2 Equivalence clustering $G_{\mathbf{Po}}$ and G_{box}

Here we prove Proposition 4.3.4. We first establish a few results regarding the number of nodes of degree k_n in both the Poissonized KPKVB graph G_{Po} and the finite box model G_{box}.

Lemma 4.7.3. *Let $\alpha > 1/2$, $\nu > 0$ and $(k_n)_{n\geq 1}$ be a sequence such that $k_n = O\left(n^{1/(2\alpha+1)}\right)$. Then*

$$\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right] = \Theta(1)\, n k_n^{-(2\alpha+1)}, \tag{4.83}$$

and

$$\mathbb{E}\left[N_{\mathrm{box}}(k_n)\right] = \Theta(1)\, n k_n^{-(2\alpha+1)}. \tag{4.84}$$

Moreover,

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]}{\mathbb{E}\left[N_{\mathrm{box}}(k_n)\right]} = 1. \tag{4.85}$$

Proof. By the Campbell-Mecke formula

$$\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right] = \int_{\mathcal{R}} \rho_{\mathrm{Po}}(y, k_n) f(x,y)\, \mathrm{d}x\, \mathrm{d}y.$$

Then by Lemma 4.4.7

$$\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right] = (1 + o(1)) \int_{\mathcal{R}} \rho(y, k_n) f(x,y)\, \mathrm{d}x\, \mathrm{d}y$$

$$= (1 + o(1))n \int_0^R \rho(y, k_n) f(x, y) \, dx \, dy = \Theta(1) \, n k_n^{-(2\alpha+1)}.$$

Similarly,

$$\mathbb{E}\left[N_{\text{box}}(k_n)\right] = (1 + o(1)) \int_{\mathcal{R}} \rho(y, k_n) f(x, y) \, dx \, dy,$$

from which the results follow. □

Recall that Proposition 4.3.4 states that

$$\lim_{n\to\infty} s(k_n)^{-1} \, \mathbb{E}\left[|c^*(k_n; G_{\text{Po}}) - c^*(k_n; G_{\text{box}})|\right] = 0.$$

Next recall the definition of $\mathcal{K}_C(k_n)$ as

$$\mathcal{K}_C(k_n) = \left\{ y \in \mathbb{R}_+ : \frac{k_n - C\sqrt{k_n \log(k_n)}}{\xi} \vee 0 \le e^{\frac{y}{2}} \le \frac{k_n + C\sqrt{k_n \log(k_n)}}{\xi} \wedge e^{R/2} \right\},$$

and (4.63)

$$\widetilde{c}_{\text{box}}(k_n) = \frac{\widetilde{T}_{\text{box}}(k_n, C)}{\binom{k_n}{2} \mathbb{E}\left[N_{\text{box}}(k_n)\right]},$$

where $\widetilde{T}_{\text{box}}(k_n, C)$ counts for all nodes $p = (x, y)$ with $y \in \mathcal{K}_C(k_n)$ the pairs (p_1, p_2) that form a triangle with p, with the exception that it considers $p_2 \in \mathcal{B}_\infty(p_1) \cap \mathcal{R}$ instead of $\mathcal{B}_{\text{box}}(p_1)$. Then using Corollary 4.6.2 we get

$$\mathbb{E}\left[|c^*(k_n; G_{\text{Po}}) - c^*(k_n; G_{\text{box}})|\right] \le \mathbb{E}\left[|c^*(k_n; G_{\text{Po}}) - \widetilde{c}_{\text{box}}(k_n)|\right] + o(s(k_n)),$$

and hence it is enough to prove that

$$\lim_{n\to\infty} s(k_n)^{-1} \mathbb{E}\left[|c^*(k_n; G_{\text{Po}}) - \widetilde{c}_{\text{box}}(k_n)|\right] = 0.$$

The following lemma will be frequently used in the proof of Proposition 4.3.4.

Lemma 4.7.4. *Let $t, r \in \mathbb{R}$ be fixed and let $\hat{\rho}(y, k)$ be any of the three probability functions $\rho_{\text{Po}}(y, k), \rho_{box}(y, k)$ or $\rho(y, k)$. Then for any sequence k_n of non-negative integers with $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$ and $C > 0$ large enough,*

$$\int_{\mathcal{K}_C} e^{ty} \hat{\rho}_n(y, k_n - r) e^{-\alpha y} \, dy = O(1) \, k_n^{-2\alpha-1+2t}$$

as $n \to \infty$.

Proof. Note that on $\mathcal{K}_C(k_n)$ we have that $e^{ty} = \Theta\left(k_n^{2t}\right)$. Hence, by the second statement of Lemma 4.4.7

$$\begin{aligned}\int_{\mathcal{K}_C} e^{ty} \hat{\rho}_n(y, k_n - r) e^{-\alpha y} \, dy &= \Theta\left(k_n^{2t}\right) \int_{\mathcal{K}_C} \hat{\rho}_n(y, k_n - r) e^{-\alpha y} \, dy \\ &= O\left(k_n^{2t}\right)(k_n - r)^{-(2\alpha+1)} = O(1) \, k_n^{-2\alpha-1+2t}.\end{aligned}$$

□

Proof of Proposition 4.3.4. To keep notation concise we abbreviate $\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]$ and $\mathbb{E}\left[N_{\mathrm{box}}(k_n)\right]$ by $\overline{n}_{\mathrm{Po}}(k_n)$ and $\overline{n}_{\mathrm{box}}(k_n)$, respectively. We will also suppress the subscript n in most expressions regarding the graphs G_{Po} and G_{box}. Finally we will write

$$T_{\mathrm{Po}}(p) = \sum_{p_1,p_2\in\mathcal{P}\setminus\{p\}}^{\neq} T_{\mathrm{Po}}(p,p_1,p_2),$$

with

$$T_{\mathrm{Po}}(p,p_1,p_2) = \mathbb{1}_{\{p_1\in\mathcal{B}(p)\}}\mathbb{1}_{\{p_2\in\mathcal{B}(p)\}}\mathbb{1}_{\{p_2\in\mathcal{B}(p_1)\}}$$

to denote the triangle count function for p in G_{Po}. Then we have

$$\begin{aligned}
&\mathbb{E}\left[\left|c^*(k_n;G_{\mathrm{Po}}) - \widetilde{c}_{\mathrm{box}}(k_n)\right|\right]\\
&= \binom{k_n}{2}^{-1}\mathbb{E}\left[\left|\sum_{p\in\mathcal{P}}\frac{\mathbb{1}_{\{\deg_{\mathrm{Po}}(p)=k_n\}}}{\overline{n}_{\mathrm{Po}}(k_n)}T_{\mathrm{Po}}(p) - \frac{\mathbb{1}_{\{\deg_{\mathrm{box}}(p)=k_n\}}}{\overline{n}_{\mathrm{box}}(k_n)}\widetilde{T}_{\mathrm{box}}(p)\right|\right]\\
&\le \binom{k_n}{2}^{-1}\overline{n}_{\mathrm{Po}}(k_n)^{-1}\mathbb{E}\left[\left|\sum_{p\in\mathcal{P}}\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}}T_{\mathrm{Po}}(p) - \mathbb{1}_{\{\deg_{\mathrm{box}}(p)=k_n\}}\widetilde{T}_{\mathrm{box}}(p)\right|\right]\\
&\quad + \binom{k_n}{2}^{-1}\left|\frac{1}{\overline{n}_{\mathrm{Po}}(k_n)} - \frac{1}{\overline{n}_{\mathrm{box}}(k_n)}\right|\mathbb{E}\left[\sum_{p\in\mathcal{P}}\mathbb{1}_{\{\deg_{\mathrm{box}}(p)=k_n\}}\widetilde{T}_{\mathrm{box}}(p)\right].
\end{aligned}$$

The last term can be rewritten as

$$\left|1-\frac{\overline{n}_{\mathrm{Po}}(k_n)}{\overline{n}_{\mathrm{box}}(k_n)}\right|\mathbb{E}\left[\widetilde{c}_{\mathrm{box}}(k_n)\right] = \left|1-\frac{\overline{n}_{\mathrm{Po}}(k_n)}{\overline{n}_{\mathrm{box}}(k_n)}\right|\gamma(k_n)(1+o(1)),$$

where we have used Proposition 4.3.6 (See Section 4.5). The first term in this product converges to zero by Lemma 4.7.3 while the second term scales as $s(k_n)$. Hence

$$\left|1-\frac{\overline{n}_{\mathrm{Po}}(k_n)}{\overline{n}_{\mathrm{box}}(k_n)}\right|\mathbb{E}\left[\widetilde{c}_{\mathrm{box}}(k_n)\right] = o\left(s(k_n)\right),$$

and therefore we are left to analyze the other term. By the Campbell-Mecke formula

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{p\in\mathcal{P}}\mathbb{1}_{\{\deg_{\mathrm{Po}}(p)=k_n\}}T_{\mathrm{Po}}(p) - \mathbb{1}_{\{\deg_{\mathrm{box}}(p)=k_n\}}\widetilde{T}_{\mathrm{box}}(p)\right|\right]\\
&= \int_{\mathcal{R}}\mathbb{E}\left[\left|\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}}T_{\mathrm{Po}}(y) - \mathbb{1}_{\{\deg_{\mathrm{box}}(y)=k_n\}}\widetilde{T}_{\mathrm{box}}(y)\right|\right]f(x,y)\,\mathrm{d}y\,\mathrm{d}x.
\end{aligned}$$

Since

$$\mathbb{E}\left[\frac{\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}}}{\overline{n}_{\mathrm{Po}}(k_n)}T_{\mathrm{Po}}(y)\right] \le \binom{k_n}{2}\rho_{\mathrm{Po}}(y,k_n)\overline{n}_{\mathrm{Po}}(k_n)^{-1}$$

$$= \binom{k_n}{2} \rho_{\mathrm{Po}}(y, k_n) \Theta\left(\overline{n}_{\mathrm{box}}(k_n)^{-1}\right)$$
$$= \Theta\left(n^{-1} k_n^{2\alpha+3}\right) \rho_{\mathrm{Po}}(y, k_n),$$

and similarly for the other term, it follows that

$$\mathbb{E}\left[\left|\frac{\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}}}{\overline{n}_{\mathrm{Po}}(k_n)} T_{\mathrm{Po}}(y) - \frac{\mathbb{1}_{\{\deg_{\mathrm{box}}(y)=k_n\}}}{\overline{n}_{\mathrm{Po}}(k_n)} \widetilde{T}_{\mathrm{box}}(y)\right|\right]$$
$$\leq \Theta\left(n^{-1} k_n^{2\alpha+3}\right) \left(\rho_{\mathrm{Po}}(y, k_n) + \rho_{\mathrm{box}}(y, k_n)\right).$$

Therefore, by a concentration of heights argument (c.f. first statement of Lemma 4.4.7), it is enough to consider the integral

$$n \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[\left|\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}} T_{\mathrm{Po}}(y) - \mathbb{1}_{\{\deg_{\mathrm{box}}(y)=k_n\}} \widetilde{T}_{\mathrm{box}}(y)\right|\right] e^{-\alpha y} \, \mathrm{d}y, \tag{4.86}$$

where we have also used that $f(x, y)$ is simply a constant multiple of the function $e^{-\alpha y}$. Since $\binom{k_n}{2}\overline{n}_{\mathrm{Po}}(k_n) = \Theta\left(n k_n^{-(2\alpha-1)}\right)$ we have to show that

$$\lim_{n\to\infty} k_n^{2\alpha-1} s(k_n)^{-1}$$
$$\times \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[\left|\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}} T_{\mathrm{Po}}(y) - \mathbb{1}_{\{\deg_{\infty}(y)=k_n\}} \widetilde{T}_{\mathrm{box}}(y)\right|\right] e^{-\alpha y} \, \mathrm{d}y = 0.$$

For $\alpha > 3/4$, $s_{3/4}(k_n) = \log(k_n)^{-1} s_\alpha(k_n) = o\left(s_\alpha(k_n)\right)$ and thus it suffices to prove the following two cases:

1. if $1/2 < \alpha \leq 3/4$, then

$$\lim_{n\to\infty} k_n^{6\alpha-3} \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[\left|\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}} T_{\mathrm{Po}}(y) - \mathbb{1}_{\{\deg_{\mathrm{box}}(y)=k_n\}} \widetilde{T}_{\mathrm{box}}(y)\right|\right] e^{-\alpha y} \, \mathrm{d}y = 0,$$

2. if $3/4 < \alpha$, then

$$\lim_{n\to\infty} k_n^{2\alpha} \int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[\left|\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}} T_{\mathrm{Po}}(y) - \mathbb{1}_{\{\deg_{\mathrm{box}}(y)=k_n\}} \widetilde{T}_{\mathrm{box}}(y)\right|\right] e^{-\alpha y} \, \mathrm{d}y = 0.$$

We shall proceed by expanding the integrand and analyzing the individual terms. With a slight abuse of notation we shall write y instead of $(0, y)$ in an expression such as $\mathcal{B}(y)$. In addition we write $D_{\mathrm{Po}}(y, k_n; \mathcal{P})$ for the indicator that is equal to 1 if and only if $\mathcal{B}(y)$ contains k_n points from $\mathcal{P} \setminus \{(0, y)\}$. We define $D_{\mathrm{box}}(y, k_n; \mathcal{P})$ analogously for the ball $\mathcal{B}_{\mathrm{box}}(y)$. It is important to note that for any $p' \in \mathcal{R}$ it holds that $p' \in \mathcal{B}_{\mathrm{box}}(y) \iff p' \in \mathcal{B}_\infty(y)$.

We need to split the integrand over several terms and then analyze each of these separately. Applying the Campbell-Mecke formula yields

$$\mathbb{E}\left[\left|\mathbb{1}_{\{\deg_{\mathrm{Po}}(y)=k_n\}} P_{\mathrm{Po}}(y) - \mathbb{1}_{\{\deg_{\infty}(y)=k_n\}} \widetilde{T}_{\mathrm{box}}(y)\right|\right] \leq$$

$$\mathbb{E}\left[\sum_{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}}^{\neq} \Big| D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})T_{\mathrm{Po}}(y,p_1,p_2) - D_{\mathrm{box}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\widetilde{T}_{\mathrm{box}}(y,p_1,p_2)\Big|\right],$$

where the sum ranges over all distinct pairs of points in $\mathcal{P}\setminus\{(0,y)\}$. In what follows, we will set $\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p') = \mathcal{B}(p')\triangle(\mathcal{B}_\infty(p')\cap\mathcal{R})$ and $\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(p') = \mathcal{B}(p')\cap\mathcal{B}_{\mathrm{box}}(p')$ and observe that $\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y) = \mathcal{B}(y)\cap\mathcal{B}_\infty(y)$. We will now bound the sum that is inside the expectation. We will split the sum into different parts, depending on combinations of $p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}$ for which only one of the two terms of the difference is non-zero. Clearly, for this we need that either $p_1\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y)$ and $p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)$, or $p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)$ and $p_2\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(p_1)$. We will consider the following four cases:

1. $p_1\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y)$ and $p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)$ and
   a) $y_1,y_2<(1-\varepsilon)R\wedge(R-y)$,
   b) $y_1\geq(1-\varepsilon)R\wedge(R-y)$,
2. $p_1\in\mathcal{B}(y)\setminus\mathcal{B}_\infty(y)$ with $y_1<K$ and $p_2\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y)$,
3. $p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)$ with $y_1\geq K$ and $p_2\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y)$,

where K in the last two cases is the constant from Lemma 1.6.2.

Observe that when $y_1<(1-\varepsilon)R\wedge(R-y)$ and $y_2\geq(1-\varepsilon)R\wedge(R-y)$ it follows from Corollary 4.7.1 that $p_2\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(p_1)$ and thus we do not have to consider this case when $p_1\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y)$ and $p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)$. Similarly, when $y_1\geq K$ and $p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)$ Corollary 4.7.1 implies that $p_1\in\mathcal{B}(y)\setminus\mathcal{B}_\infty(y)$ which explains the setting of case 2.

We can now bound the sum by the following expression:

$$\sum_{(p_1,p_2)\in\mathcal{P}\setminus\{(0,y)\}}^{\neq} \Big| D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})T_{\mathrm{Po}}(y,p_1,p_2) - D_{\mathrm{box}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\widetilde{T}_{\mathrm{box}}(y,p_1,p_2)\Big|$$

$$\leq \sum_{\substack{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}\\ y_1,y_2<(1-\varepsilon)R\wedge(R-y)}}^{\neq} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\}}\,D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\}) \tag{4.87}$$

$$+ \sum_{\substack{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}\\ y_1,y_2<(1-\varepsilon)R\wedge(R-y)}}^{\neq} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\cap\mathrm{box}}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\}}\,D_{\mathrm{box}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\}) \tag{4.88}$$

$$+ \sum_{\substack{p_1,p_2 \ \in \mathcal{P}\setminus\{(0,y)\} \\ y_1 \geq (1-\varepsilon)R \wedge (R-y)}}^{\neq} \mathbb{1}_{\{p_1 \in \mathcal{B}_{\mathrm{Po}} \cap \mathrm{box}(y)\}} \cdot \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)\}} D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\}) \tag{4.89}$$

$$+ \sum_{\substack{p_1,p_2 \ \in \mathcal{P}\setminus\{(0,y)\} \\ y_1 \geq (1-\varepsilon)R \wedge (R-y)}}^{\neq} \mathbb{1}_{\{p_1 \in \mathcal{B}_{\mathrm{Po}} \cap \mathrm{box}(y)\}} \cdot \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)\}} D_{\mathrm{box}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\}) \tag{4.90}$$

$$+ \sum_{\substack{p_1,p_2 \in \mathcal{P}\setminus\{(0,y)\} \\ y(p_1) \geq K}}^{\neq} \mathbb{1}_{\{p_1 \in \mathcal{B}(y) \setminus \mathcal{B}_\infty(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}(y) \cap \mathcal{B}_\infty(y)\}} D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\}) \tag{4.91}$$

$$+ \sum_{\substack{p_1,p_2 \in \mathcal{P}\setminus\{(0,y)\} \\ y(p_1) \geq K}}^{\neq} \mathbb{1}_{\{p_1 \in \mathcal{B}(y) \setminus \mathcal{B}_\infty(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}(y) \cap \mathcal{B}_\infty(y)\}} D_{\mathrm{box}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\}) \tag{4.92}$$

$$+ \sum_{\substack{p_1,p_2 \in \mathcal{P}\setminus\{(0,y)\} \\ y(p_1) < K}}^{\neq} \mathbb{1}_{\{p_1 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}(y) \cap \mathcal{B}_\infty(y)\}}. \tag{4.93}$$

In the following paragraphs we will give upper bounds on the expected values of each one of these partial sums.

The sums (4.87) **and** (4.88) We will analyze (4.87). The analysis of the other sum (4.88) is similar. Note first that for any two points p_1, p_2 the following holds: if $p_1 \in \mathcal{B}(y)$ and $p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)$, then $p_2 \in \mathcal{B}(y)$ and $p_1 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_2) \cap \mathcal{B}(y)$. Using this symmetry, it suffices to consider distinct pairs $(p_1, p_2) \in \mathcal{P} \setminus \{(0, y)\}$ with $0 \leq y_2 \leq y_1 \leq R - y$. Let $\mathcal{D}$ denote the set of these pairs.

We are going to consider several sub-cases and, thereby, split the domain $\mathcal{D}$ into the corresponding sub-domains. Let $\omega = \omega(n) \to \infty$ as $n \to \infty$ be a slowly growing function and set $y_\omega := y + \omega$. We let

$$\begin{aligned}
\mathcal{D}_1 &= \{(p_1, p_2) \in \mathcal{D} \cap \mathcal{P} \ : \ y \leq y_1 \leq R/2,\, y_\omega \leq y_2 \leq y_1\}, \\
\mathcal{D}_2 &= \{(p_1, p_2) \in \mathcal{D} \cap \mathcal{P} \ : \ y_1 \leq R/2,\, y_2 \leq y_\omega\} \text{ and} \\
\mathcal{D}_3 &= \{(p_1, p_2) \in \mathcal{D} \cap \mathcal{P} \ : R/2 < y_1 \leq R - y,\, y_2 \leq y_1\}.
\end{aligned}$$

Note that $\mathcal{D} \subseteq \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3$. Hence, we can write

$$\mathbb{E}\left[\sum_{\substack{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}\\ y_1,y_2\le(1-\varepsilon)R\wedge(R-y)}} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}\, D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right]$$
$$\le \sum_{i=1}^{3}\mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_i} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}\cdot D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right]. \tag{4.94}$$

We bound each one of the above three summands as

$$\mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_1} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}\, D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right]$$
$$\le \mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_1} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}(y)\}}\, D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right] := \mathcal{I}_n^{(1)}(y), \tag{4.95}$$

$$\mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_2} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}\, D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right]$$
$$\le \mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_2} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\}}\, D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right] := \mathcal{I}_n^{(2)}(y), \tag{4.96}$$

and

$$\mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_3} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}\, D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right]$$
$$\le \mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_3} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\cdot\mathbb{1}_{\{p_2\in\mathcal{B}(y)\}}\, D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right] := \mathcal{I}_n^{(3)}(y). \tag{4.97}$$

We will bound each term using the Campbell-Mecke formula and show for $i = 1,2,3$ that, for $1/2 < \alpha < 3/4$,

$$\lim_{n\to\infty} k_n^{6\alpha-3}\int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(i)}(y)e^{-\alpha}\,\mathrm{d}y = 0, \tag{4.98}$$

and, for $\alpha \geq 3/4$,

$$\lim_{n\to\infty} k_n^{2\alpha} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(i)}(y) e^{-\alpha}\, \mathrm{d}y = 0. \tag{4.99}$$

For the first term (4.95), we note that

$$\mathbb{E}\left[D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\}\right] = \rho_{\mathrm{Po}}(y, k_n - 2).$$

and hence $\mathcal{I}_n^{(1)}(y)$ becomes

$$\begin{aligned} \rho_{\mathrm{Po}}(y, k_n - 2) \int_{-I_n}^{I_n} \int_{y}^{R/2} \int_{-I_n}^{I_n} \int_{y_\omega}^{y_1} & \mathbb{1}_{\{p_1 \in \mathcal{B}_{\mathrm{Po}} \cap \mathrm{box}(y)\}} \\ & \times \mathbb{1}_{\{p_2 \in \mathcal{B}(y)\}} e^{-\alpha(y_1+y_2)}\, \mathrm{d}y_2\, \mathrm{d}x_2\, \mathrm{d}y_1\, \mathrm{d}x_1. \end{aligned} \tag{4.100}$$

Next, Lemma 1.6.2 implies that for $y' \leq R - y$, we have that if $(x', y') \in \mathcal{B}(y)$, then $|x'| < (1+K)e^{y/2+y'/2}$, where $K > 0$ is as in Lemma 1.6.2. Using these observations, we obtain

$$\begin{aligned} & \mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_1} \mathbb{1}_{\{p_1 \in \mathcal{B}_{\mathrm{Po}} \cap \mathrm{box}((0,y))\}} \cdot \mathbb{1}_{\{p_2 \in \mathcal{B}(y)\}} \cdot D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\})\right] \\ & \quad = \rho_{\mathrm{Po}}(y, k_n - 2) e^{y} \int_{y}^{R/2} e^{y_1/2} \int_{y_\omega}^{y_1} e^{y_2/2} e^{-\alpha y_2} \cdot e^{-\alpha y_1} dy_2 dy_1. \end{aligned}$$

Now, the double integral becomes

$$\begin{aligned} & \int_{y}^{R/2} e^{y_1/2} \int_{y_\omega}^{y_1} e^{y_2/2} e^{-\alpha y_2} \cdot e^{-\alpha y_1} dy_2 dy_1 = \\ & O(1) \cdot \int_{y}^{R/2} e^{y_1/2 - \alpha y_1} \cdot e^{(1/2-\alpha)y_\omega} dy_1 \\ & = O(1) \cdot e^{(1/2-\alpha)y_\omega} \cdot \int_{y}^{R/2} e^{y_1/2 - \alpha y_1} dy_1 \\ & = O(1) \cdot e^{(1/2-\alpha)y_\omega + (1/2-\alpha)y} \\ & \ll e^{(1-2\alpha)y}, \end{aligned} \tag{4.101}$$

since $y_\omega = y + \omega$ and $\omega \to \infty$. We then deduce that

$$\begin{aligned} & \mathbb{E}\left[\sum_{p_1,p_2\in\mathcal{D}_1} \mathbb{1}_{\{p_1 \in \mathcal{B}_{\mathrm{Po}} \cap \mathrm{box}((0,y))\}} \cdot \mathbb{1}_{\{p_2 \in \mathcal{B}(y)\}} \cdot D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\})\right] \\ & \ll \rho_{\mathrm{Po}}(y, k_n - 2) e^{(1-2\alpha)y}. \end{aligned} \tag{4.102}$$

We now integrate this with respect to y and determine its contribution to (4.86),

$$\begin{aligned}&\int_{\mathcal{K}_C(k_n)} \rho_{\mathrm{Po}}(y, k_n - 2) e^{(1-2\alpha)y} e^{-\alpha y} \, dy \, dx \\ &= O\left(k_n^{-6\alpha+1}\right),\end{aligned}$$

where we have used Lemma 4.7.4 with $t = 1 - 2\alpha$.

Since $1 - 6\alpha + \min\{6\alpha - 3, 2\alpha\} < 0$ for all $\alpha > 1/2$ we deduce that, for $1/2 < \alpha < 3/4$,

$$\lim_{n \to \infty} k_n^{6\alpha-3} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(1)}(y) e^{-\alpha y} \, dy = 0,$$

while, for $\alpha \geq 3/4$,

$$\lim_{n \to \infty} k_n^{2\alpha} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(1)}(y) e^{-\alpha y} \, dy = 0.$$

We will now bound the term in (4.96). Using similar observations as for the previous term we get that $\mathcal{I}_n^{(2)}(y)$ equals

$$\begin{aligned}\rho_{\mathrm{Po}}(y, k_n - 2) \int_{-I_n}^{I_n} \int_0^{R/2} \int_{-I_n}^{I_n} \int_0^{y_\omega} &\mathbb{1}_{\{p_1 \in \mathcal{B}(y)\}} \\ &\times \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}((0,y))\}} e^{-\alpha(y_1+y_2)} \, dy_2 \, dx_2 \, dy_1 \, dx_1.\end{aligned}$$

Now, Lemma 1.6.2 implies that for $y_2 \leq R - y_1$, we have that if $(x_2, y_2) \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}((x_1, y_1))$, then x_2 lies in an interval of length $Ke^{3y/2+3y'/2-R}$, where $K > 0$ is again the constant in Lemma 1.6.2. Using these observations we obtain

$$\mathcal{I}_n^{(2)}(y) = \rho_{\mathrm{Po}}(y, k_n - 2) e^{y/2} \int_0^{R/2} e^{y_1/2+3y_1/2} \int_0^{y_\omega} e^{3y_2/2-R} e^{-\alpha y_2} \cdot e^{-\alpha y_1} \, dy_2 \, dy_1. \tag{4.103}$$

The integrals satisfy

$$\begin{aligned}&e^{-R}\left(\int_0^{R/2} e^{(2-\alpha)y_1} \, dy_1\right)\left(\int_0^{y_\omega} e^{(3/2-\alpha)y_2} \, dy_2\right) \\ &= O(1)\, e^{-R}\left(\begin{cases} e^{(1-\alpha/2)R} & \text{if } \frac{1}{2} < \alpha < 2 \\ R & \text{if } \alpha \geq 2 \end{cases}\right)\left(\begin{cases} e^{(3/2-\alpha)y_\omega} & \text{if } \frac{1}{2} < \alpha < \frac{3}{2} \\ y & \text{if } \alpha \geq \frac{3}{2} \end{cases}\right) \\ &= O(1) \begin{cases} e^{-\frac{\alpha}{2}R} e^{(3/2-\alpha)y} & \text{if } \frac{1}{2} < \alpha < \frac{3}{2}, \\ (y + \omega(n)) e^{-\frac{\alpha}{2}R} & \text{if } \frac{3}{2} \leq \alpha < 2, \\ (y + \omega(n)) R e^{-R} & \text{if } \alpha \geq 2. \end{cases}\end{aligned}$$

Since $y_\omega := y + \omega(n) \leq R = O(\log(n))$ we conclude that on $\mathcal{K}_C(k_n)$

$$\mathcal{I}_n^{(2)}(y) = O(1)\, \rho_{\mathrm{Po}}(y, k_n - 2) \begin{cases} n^{-\alpha} k_n^{3-2\alpha} & \text{if } \frac{1}{2} < \alpha < \frac{3}{2}, \\ n^{-\alpha} \log(n) & \text{if } \frac{3}{2} \leq \alpha < 2, \\ n^{-2} \log(n)^2 & \text{if } \alpha \geq 2, \end{cases}$$

and hence

$$\int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(2)}(y)e^{-\alpha y}\,\mathrm{d}y = O\left(1\right)k_n^{-(2\alpha+1)}\begin{cases} n^{-\alpha}k_n^{3-2\alpha} & \text{if } \frac{1}{2}<\alpha<\frac{3}{2}, \\ n^{-\alpha}\log(n) & \text{if } \frac{3}{2}\leq\alpha<2, \\ n^{-2}\log(n)^2 & \text{if } \alpha\geq 2 \end{cases}$$
$$= O\left(1\right)\begin{cases} n^{-\alpha}k_n^{2-4\alpha} & \text{if } \frac{1}{2}<\alpha<\frac{3}{2}, \\ n^{-\alpha}\log(n)k_n^{-(2\alpha+1)} & \text{if } \frac{3}{2}\leq\alpha<2, \\ n^{-2}\log(n)^2k_n^{-(2\alpha+1)} & \text{if } \alpha\geq 2. \end{cases}$$

Now $4\alpha^2-\alpha+1>0$ for $1/2<\alpha<3/4$. Hence since $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right)$, we have

$$k_n^{6\alpha-3}n^{-\alpha}k_n^{2-4\alpha} = n^{-\alpha}k_n^{2\alpha-1} = O\left(n^{-\alpha+\frac{2\alpha-1}{2\alpha+1}}\right) = O\left(k_n^{-\frac{4\alpha^2-\alpha+1}{2\alpha+1}}\right) = o\left(1\right),$$

from which we deduce that

$$\lim_{n\to\infty} k_n^{6\alpha-3}\int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(2)}(y)e^{-\alpha y}\,\mathrm{d}y = 0.$$

For $\alpha\geq 3/4$ we have that both $n^{-\alpha}\log(n)k_n^{-1}$ and $n^{-2}\log(n)^2k_n^{-1}$ converge to zero as $n\to\infty$ and hence in this case

$$\lim_{n\to\infty} k_n^{2\alpha}\int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(2)}(y)e^{-\alpha y}\,\mathrm{d}y = 0.$$

We will now consider the term in (4.97). Recall that $\mathcal{D}_3$ consists of all pairs $(p_1,p_2)\in\mathcal{D}$ such that $R/2<y_1\leq(1-\varepsilon)R\wedge(R-y)$ and $y_1\leq y_\omega$ with the property that $p_1\in\mathcal{B}\left(y\right)$ and $p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}\left(p_1\right)\cap\mathcal{B}\left(y\right)$. So, in particular, $p_2\in\left(\mathcal{B}\left(p_1\right)\cup\mathcal{B}_\infty\left(p_1\right)\right)\cap\mathcal{B}\left(y\right)$.

We will consider this intersection more closely. We use Lemma 1.6.2 to define a ball around p_1 that contains both $\mathcal{B}\left(p_1\right)$ and $\mathcal{B}_\infty\left(p_1\right)$. For $K>0$, we define, for any point $p_1=(x_1,y_1)\in\mathbb{R}\times\mathbb{R}_+$,

$$\check{\mathcal{B}}_{\mathrm{Po}}(p_1) := \{(x',y')\ :\ y'<R-y_1,\ |x_1-x'|<(1+K)e^{\frac{1}{2}(y_1+y')}\}. \tag{4.104}$$

It is an implication of Lemma 1.6.2 that

$$\left(\mathcal{B}\left(p_1\right)\cup\mathcal{B}_\infty\left(p_1\right)\right)\cap\mathcal{R}([0,R-y_1])\subseteq\check{\mathcal{B}}_{\mathrm{Po}}(p_1).$$

Therefore, any point $p_2=(x_2,y_2)\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}\left(p_1\right)\cap\mathcal{B}\left(y\right)$ with $y_2\leq R-y_1$ must belong to $\check{\mathcal{B}}_{\mathrm{Po}}(p_1)\cap\check{\mathcal{B}}_{\mathrm{Po}}(y)$.

We will use this in order to derive a lower bound on y_2 as a function of x_1,y_1. Let us suppose without loss of generality that $x_1<0$. The left boundary of

$\check{\mathcal{B}}_{\mathrm{Po}}((0,y))$ is given by the equation $x' = (1-K)e^{\frac{1}{2}(y+y')}$ whereas the right boundary of $\check{\mathcal{B}}_{\mathrm{Po}}(p_1)$ is given by the curve having equation $x' = x_1 + (1+K)e^{\frac{1}{2}(y_1+y')}$. The equation that determines the intersection point $(\hat{x},\hat{y})$ of these curves is

$$x_1 + (1+K)e^{(y_1+\hat{y})/2} = (1-K)e^{(y+\hat{y})/2}.$$

We can solve the above for $\hat{y}$ as

$$|x_1| = (1+K)e^{\hat{y}/2}\left(e^{y_1/2} + e^{y/2}\right).$$

But $y_1 > R/2$ and since $y \in \mathcal{K}_C(k_n)$, it follows that for sufficiently large n, $y \leq (1+\varepsilon)R/(2\alpha+1)$. So if ε is small enough depending on α, we have

$$|x_1| = (1+K)e^{\hat{y}/2}\left(e^{y_1/2} + e^{y/2}\right) = (1+K+o(1))e^{\hat{y}/2+y_1/2}.$$

Let c_K^2 denote the multiplicative term $1+K+o(1)$, which appears in the above. The above yields

$$\hat{y} = \left(2\log(|x_1|e^{-y_1/2}) - \log c_K\right) \vee 0 := \hat{y}(x_1,y_1). \tag{4.105}$$

In particular, note that $\hat{y} = 0$ if and only if $|x_1| \leq c_K e^{y_1/2}$. Moreover, since $p_1 \in \mathcal{B}(y)$ and $x_1 \leq R - y$, we also have that $|x_1| \leq e^{(y+y_1)/2}(1+o(1))$. This upper bound on $|x_1|$ together with (4.105), implies that for n sufficiently large, we have $\hat{y} \leq y$. This observation will be used below, where we integrate over y_2, thus ensuring that the integrals are non-zero.

We conclude that

$$p' \in \check{\mathcal{B}}_{\mathrm{Po}}(y) \cap \check{\mathcal{B}}_{\mathrm{Po}}((x_1,y_1)) \Rightarrow y' \geq \hat{y}(x_1,y_1),$$

which implies

$$\mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}} \leq \mathbb{1}_{\left\{y_2 \geq \hat{y}(x_1,y_1),\, p_2 \in \check{\mathcal{B}}_{\mathrm{Po}}((0,y))\right\}}. \tag{4.106}$$

If we integrate this over x_2, y_2, then we get

$$\begin{aligned}
&\int_{-I_n}^{I_n}\int_0^{y_1} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}} e^{-\alpha y_2}\, dy_2 dx_2 \\
&\leq \int_{-I_n}^{I_n}\int_0^{y_1} \mathbb{1}_{\left\{y_2 \geq \hat{y}(x_1,y_1),\, p_2 \in \check{\mathcal{B}}_{\mathrm{Po}}(y)\right\}} e^{-\alpha y_2}\, dy_2 dx_2 \\
&\leq (1+K)\cdot e^{y/2}\int_{\hat{y}(x_1,y_1)}^{y_1} e^{y_2/2-\alpha y_2}\, dy_2 \\
&= O(1)\cdot e^{y/2+(1/2-\alpha)\hat{y}(x_1,y_1)}.
\end{aligned}$$

Note also that

$$\mathbb{E}\left[D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right]=\rho_{\mathrm{Po}}(y,k_n-2),$$

uniformly over all $(p_1,p_2)\in\mathcal{D}_3$. Hence the Campbell-Mecke formula yields that $\mathcal{I}_n^{(3)}(y)$ equals

$$O(1)\rho_{\mathrm{Po}}(y,k_n-2)\,e^{y/2}\int_{-I_n}^{I_n}\int_{R/2}^{(R-y)\wedge(1-\varepsilon)R}\mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}e^{(1/2-\alpha)\hat{y}(x_1,y_1)-\alpha y_1}dy_1dx_1$$
$$=O(1)\rho_{\mathrm{Po}}(y,k_n-2)\,e^{y/2}\int_{-I_n}^{I_n}\int_{R/2}^{(R-y)\wedge(1-\varepsilon)R}\mathbb{1}_{\left\{p_1\in\check{\mathcal{B}}_{\mathrm{Po}}(y)\right\}}e^{(1/2-\alpha)\hat{y}(x_1,y_1)-\alpha y_1}dy_1dx_1.$$

Due to the symmetry of $\check{\mathcal{B}}_{\mathrm{Po}}(y)$, the integration over x_1 is

$$O(1)\cdot e^{y/2}\cdot\int_0^{(1+K)e^{y/2+y_1/2}}e^{\hat{y}(x_1,y_1)(1/2-\alpha)}dx_1$$

We will split this integral into two parts according to the value of $\hat{y}(x_1,y_1)$ as

$$\int_0^{(1+K)e^{y/2+y_1/2}}e^{\hat{y}(x_1,y_1)(1/2-\alpha)}dx_1$$
$$=\int_{c_Ke^{y_1/2}}^{(1+K)e^{y/2+y_1/2}}e^{\hat{y}(x_1,y_1)(1/2-\alpha)}dx_1+\int_0^{c_Ke^{y_1/2}}dx_1.$$

The first integral becomes

$$\int_{c_Ke^{y_1/2}}^{(1+K)e^{y/2+y_1/2}}e^{\hat{y}(x_1,y_1)(1/2-\alpha)}dx_1=\int_{c_Ke^{y_1/2}}^{(1+K)e^{y/2+y_1/2}}e^{\hat{y}(x_1,y_1)/2(1-2\alpha)}dx_1$$
$$=O(1)\cdot\int_{c_Ke^{y_1/2}}^{(1+K)e^{y/2+y_1/2}}x_1^{1-2\alpha}e^{-\frac{y_1}{2}(1-2\alpha)}dx_1$$
$$=O(1)\cdot e^{-y_1/2+\alpha y_1}\cdot e^{\frac{(y+y_1)}{2}2(1-\alpha)}$$
$$=O(1)\cdot e^{y_1/2+y(1-\alpha)}.$$

The second integral trivially gives

$$\int_0^{c_Ke^{y_1/2}}dx_1=O(1)\cdot e^{y_1/2}=O(1)\cdot e^{y_1/2+y(1-\alpha)}.$$

We conclude that

$$e^{y/2}\cdot\int_0^{(1+K)e^{y/2+y_1/2}}e^{\hat{y}(x_1,y_1)(1/2-\alpha)}dx_1=O(1)\cdot e^{y_1/2+y(3/2-\alpha)}.$$

Now, we integrate this with respect to y_1 and get

$$e^{y(3/2-\alpha)}\int_{R/2}^{R-y} e^{(1/2-\alpha)y_1}dy_1 = O(1)\cdot e^{y(3/2-\alpha)}e^{(1/2-\alpha)R/2} = O(1)\cdot n^{1/2-\alpha}\cdot e^{y(3/2-\alpha)},$$

from which we deduce

$$\mathcal{I}_n^{(3)}(y) = O(1)\cdot n^{1/2-\alpha}e^{y(3/2-\alpha)}\,\rho_{\mathrm{Po}}(y,k_n-2). \tag{4.107}$$

We now apply Lemma 4.7.4 with $t=\frac{3}{2}-\alpha$ and get

$$\begin{aligned}\int_{\mathcal{K}_C(k_n)}\mathcal{I}_n^{(3)}(y)e^{-\alpha y}\,\mathrm{d}y &= O\left(1\right)n^{-(\alpha-\frac{1}{2})}\int_{\mathcal{K}_C(k_n)}e^{(3/2-\alpha)y}\rho_{\mathrm{Po}}(y,k_n-2)e^{-\alpha y}\,\mathrm{d}y\\ &= O\left(n^{-(\alpha-\frac{1}{2})}k_n^{2-4\alpha}\right).\end{aligned}$$

Since for $\alpha>1/2$, $k_n = O\left(n^{\frac{1}{2\alpha+1}}\right) = o\left(n^{1/2}\right)$ we have that

$$k_n^{6\alpha-3}k_n^{2-4\alpha}n^{-(\alpha-1/2)} = o\left(1\right)$$

and hence, for $1/2<\alpha<3/4$,

$$\lim_{n\to\infty} k_n^{6\alpha-3}\int_{\mathcal{K}_C(k_n)}\mathcal{I}_n^{(3)}(y)e^{-\alpha y}\,\mathrm{d}x\,\mathrm{d}y = 0.$$

For $\alpha\geq 3/4$ we observe that $2\alpha^2+2\alpha-5/2>0$. Hence,

$$k_n^{2\alpha}n^{-(\alpha-\frac{1}{2})}k_n^{2-4\alpha} = O\left(n^{-(\alpha-1/2)}n^{\frac{2-2\alpha}{2\alpha+1}}\right) = O\left(n^{-\frac{2\alpha^2+2\alpha-5/2}{2\alpha+1}}\right) = o\left(1\right),$$

and we get, for $\alpha\geq 3/4$,

$$\lim_{n\to\infty} k_n^{2\alpha}\int_{\mathcal{K}_C(k_n)}\mathcal{I}_n^{(3)}(y)e^{-\alpha y}\,\mathrm{d}x\,\mathrm{d}y = 0.$$

The sums (4.89) and (4.90) Again, we will only consider (4.89) since the analysis for the other term is similar. Recall that in this case, we consider pairs (p_1,p_2), with $p_1=(x_1,y_1)$ satisfying $y_1\geq(R-y)\wedge(1-\varepsilon)R$, and $p_1\in\mathcal{B}\left(y\right)$, $p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}\left(p_1\right)\cap\mathcal{B}\left(y\right)$. We split this into three sub-domains: i) $y_2\geq R-y$; ii) $R-y_1\leq y_2\leq R-y$ and iii) $y_2<R-y_1$. Similarly to the analysis above we define

$$\begin{aligned}\mathcal{D}_1 &:= \{(p_1,p_2)\ :\ p_1,p_2\ \in\mathcal{P}\setminus\{(0,y)\},\ y_1\geq(1-\varepsilon)R\wedge(R-y),\ R-y\leq y_2\leq R\},\\ \mathcal{D}_2 &:= \{(p_1,p_2)\ :\ p_1,p_2\ \in\mathcal{P}\setminus\{(0,y)\},\ y_1\geq(1-\varepsilon)R\wedge(R-y),\\ &\qquad\qquad R-y_1\leq y_2\leq R-y\},\end{aligned}$$

$\mathcal{D}_3 := \{(p_1,p_2)\ :\ p_1,p_2\ \in \mathcal{P}\setminus\{(0,y)\},\ y_1 \geq (1-\varepsilon)R \wedge (R-y),\ y_2 \leq R-y_1\}.$

and write, for $i = 1,2,3$,

$$\mathcal{I}_n^{(i)}(y) := \mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_i} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\cdot \mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}\cdot D_{\mathrm{Po}}(y,k_n-2;\mathcal{P}\setminus\{p_1,p_2\})\right].$$

In the first case, note that for $y \in \mathcal{K}_C(k_n)$ we have, for small enough ε and sufficiently large n, $2y \leq 2(1+\varepsilon)\frac{R}{2\alpha+1} = o(R)$. Thus $y_1 + y_2 \geq 2(R-y) = \Omega(R)$ and thus $p_2 \in \mathcal{B}(p_1)$ for large enough n. Furthermore, $y_2 > R - y_1 + 2\ln(\pi/2)$, which implies that $p_2 \in \mathcal{B}_\infty(p_1)$ too. Hence, the contribution from these pairs is zero.

The Campbell-Mecke formula yields that

$$\mathcal{I}_n^{(1)}(y) = O(1)\int_{-I_n}^{I_n}\int_{(1-\varepsilon)R\wedge(R-y)}^{R} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\times \int_{-I_n}^{I_n}\int_{R-y}^{R} \mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}\rho_{\mathrm{Po}}(y,k_n-2)\cdot e^{-\alpha(y_2+y_1)}\,\mathrm{d}y_2\,\mathrm{d}x_2\,\mathrm{d}y_1\,\mathrm{d}x_1.$$

We proceed to bound the integral as

$$\begin{aligned}
&\int_{-I_n}^{I_n}\int_{(1-\varepsilon)R\wedge(R-y)}^{R}\mathbb{1}_{\{p_1\in\mathcal{B}(y)\}}\\
&\qquad\times\int_{-I_n}^{I_n}\int_{R-y}^{R}\mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}}e^{-\alpha(y_1+y_2)}\,\mathrm{d}y_2\,\mathrm{d}x_2\,\mathrm{d}y_1\,\mathrm{d}x_1\\
&\leq \int_{-I_n}^{I_n}\int_{(1-\varepsilon)R\wedge(R-y)}^{R}\int_{-I_n}^{I_n}\int_{R-y}^{R} e^{-\alpha(y_1+y_2)}\,\mathrm{d}y_2\,\mathrm{d}x_2\,\mathrm{d}y_1\,\mathrm{d}x_1\\
&= \left(\int_{-I_n}^{I_n}\int_{(1-\varepsilon)R\wedge(R-y)}^{R} e^{-\alpha y_1}\,\mathrm{d}y_1\,\mathrm{d}x_1\right)\left(\int_{-I_n}^{I_n}\int_{R-y}^{R} e^{-\alpha y_2}\,\mathrm{d}y_2\,\mathrm{d}x_2\right).
\end{aligned}$$

We evaluate

$$\begin{aligned}
\int_{-I_n}^{I_n}\int_{(1-\varepsilon)R\wedge(R-y)}^{R} e^{-\alpha y_1}dy_1dx_1 &= O(1)\cdot n\cdot e^{-\alpha R+((\varepsilon R)\vee y))\alpha}\\
&= O(1)\cdot n\cdot e^{-\alpha R+\alpha y+\alpha\varepsilon R},
\end{aligned}$$

and

$$\int_{-I_n}^{I_n}\int_{R-y}^{R} e^{-\alpha y_2}dy_2dx_2 = O(1)\cdot n\cdot e^{-\alpha R+\alpha y}.$$

Also, $n \cdot e^{-\alpha R} = O(1) \cdot e^{(1/2-\alpha)R}$, whereby we deduce that

$$\int_{\mathcal{D}_1} \mathbb{1}_{\{p_1 \in \mathcal{B}(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)\}} e^{-\alpha(y_1+y_2)} dy_2 dx_2 dy_1 dx_1$$
$$= O(1) \cdot e^{(1-2\alpha)R + 2\alpha y + \alpha\varepsilon R} = O(1) \cdot n^{2(1-2\alpha)+2\alpha\varepsilon} \cdot e^{2\alpha y}.$$

With these computations we obtain

$$\int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(1)}(y) e^{-\alpha y} \,\mathrm{d}x \,\mathrm{d}y = O(1) n^{2(1-2\alpha)+2\alpha\varepsilon} \int_{\mathcal{K}_C(k_n)} e^{2\alpha y} \rho_{\mathrm{Po}}(y, k_n - 2) e^{-\alpha y} \,\mathrm{d}y \,\mathrm{d}x$$
$$= O(1) n^{2(1-2\alpha)+2\alpha\varepsilon}\, k_n^{2\alpha-1}.$$

Thus, for $1/2 < \alpha < 3/4$, we have

$$k_n^{6\alpha-3}\, n^{2(1-2\alpha)+2\alpha\varepsilon}\, k_n^{2\alpha-1} = n^{2\alpha\varepsilon} \left(\frac{k_n^2}{n}\right)^{2(2\alpha-1)} = o(1),$$

provided that $\varepsilon = \varepsilon(\alpha) > 0$ is small enough, and hence, for such ε,

$$\lim_{n\to\infty} k_n^{6\alpha-3} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(1)}(y) e^{-\alpha y} \,\mathrm{d}x \,\mathrm{d}y = 0.$$

When $\alpha \geq 3/4$ we have $2(1-2\alpha) < 1/2(4\alpha - 1)$ and we get

$$k_n^{2\alpha}\, n^{2(1-2\alpha)+2\alpha\varepsilon} \cdot k_n^{2\alpha-1} \leq k_n^{4\alpha-1}\, n^{2(1-2\alpha)} n^{2\alpha\varepsilon} = o(1),$$

provided that ε is small enough, depending on α, so that

$$\lim_{n\to\infty} k_n^{2\alpha} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(1)}(y) e^{-\alpha y} \,\mathrm{d}x \,\mathrm{d}y = 0.$$

We now consider the second sub-domain $\mathcal{D}_2$. The Campbell-Mecke formula yields that

$$\mathcal{I}_n^{(2)}(y) = \mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_2} \mathbb{1}_{\{p_1 \in \mathcal{B}(y)\}} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)\}} D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1\})\right]$$
$$= O(1)\rho_{\mathrm{Po}}(y, k_n - 2) \cdot \int_{-I_n}^{I_n} \int_{(1-\varepsilon)R \wedge (R-y)}^{R} \mathbb{1}_{\{p_1 \in \mathcal{B}(y)\}} \times$$
$$\int_{-I_n}^{I_n} \int_{R-y_1}^{R-y} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)\}} e^{-\alpha(y_1+y_2)} \,\mathrm{d}y_2 \,\mathrm{d}x_2 \,\mathrm{d}y_1 \,\mathrm{d}x_1.$$

We bound the integral as

$$\int_{-I_n}^{I_n} \int_{(1-\varepsilon)R \wedge (R-y)}^{R} \mathbb{1}_{\{p_1 \in \mathcal{B}(y)\}}$$

$$\times \int_{-I_n}^{I_n} \int_{R-y_1}^{R-y} \mathbb{1}_{\{p_2 \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)\}} e^{-\alpha(y_1+y_2)}\, \mathrm{d}y_2\, \mathrm{d}x_2\, \mathrm{d}y_1\, \mathrm{d}x_1$$
$$\leq \int_{-I_n}^{I_n} \int_{(1-\varepsilon)R \wedge (R-y)}^{R} \mathbb{1}_{\{p_1 \in \mathcal{B}(y)\}} \int_{-I_n}^{I_n} \int_{R-y_1}^{R-y} \mathbb{1}_{\{p_2 \in \mathcal{B}(y)\}} e^{-\alpha(y_1+y_2)}\, \mathrm{d}y_2\, \mathrm{d}x_2\, \mathrm{d}y_1\, \mathrm{d}x_1.$$

Now, by Lemma 1.6.2,

$$\int_{-I_n}^{I_n} \int_{R-y_1}^{R-y} \mathbb{1}_{\{p_2 \in \mathcal{B}(y)\}} \cdot e^{-\alpha y_2} dy_2 dx_2 = O(1) \cdot e^{y/2} \int_{R-y_1}^{R-y} e^{(1/2-\alpha)y_2} dy_2$$
$$= O(1) \cdot e^{y/2+(1/2-\alpha)(R-y_1)}.$$

We then integrate with respect to y_1

$$\begin{aligned} O(1) \cdot e^{y/2} \cdot &\int_{-I_n}^{I_n} \int_{(1-\varepsilon)R \wedge (R-y)}^{R} \mathbb{1}_{\{p_1 \in \mathcal{B}(y)\}} e^{(1/2-\alpha)(R-y_1)} e^{-\alpha y_1} dy_1 dx_1 \\ &\leq O(1) \cdot e^{y/2+(1/2-\alpha)R} \cdot \int_{-I_n}^{I_n} \int_{(1-\varepsilon)R \wedge (R-y)}^{R} e^{(\alpha-1/2)y_1} e^{-\alpha y_1} dy_1 dx_1 \\ &= O(1) \cdot e^{y/2+(1-\alpha)R-((1-\varepsilon)R \wedge (R-y))/2} \\ &= O(1) \cdot e^{y/2+(1/2-\alpha)R+((\varepsilon R) \vee y)/2} \\ &= O(1) \cdot e^{y+(1/2-\alpha)R+\varepsilon R} = O(1) \cdot n^{1-2\alpha+\varepsilon} \cdot e^{y}. \end{aligned}$$

Therefore we get

$$\begin{aligned} &\int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(2)}(y) e^{-\alpha y}\, \mathrm{d}x\, \mathrm{d}y \\ &= O\left(n^{1-2\alpha+\varepsilon}\right) \int_{\mathcal{K}_C(k_n)} \rho_{\mathrm{Po}}(y, k_n - 2) e^{y} e^{-\alpha y}\, \mathrm{d}x\, \mathrm{d}y \\ &= O\left(1\right) n^{1-2\alpha+\varepsilon} k_n^{-2\alpha+1}, \end{aligned}$$

where we have used Lemma 4.7.4 with $t = 1$.

For $1/2 < \alpha < 3/4$, we have

$$k_n^{4\alpha-2} \cdot n^{1-2\alpha+\varepsilon} = n^{\varepsilon} \left(\frac{k_n^2}{n}\right)^{2\alpha-1} = o(1),$$

provided that $\varepsilon = \varepsilon(\alpha) > 0$ is small enough, yielding

$$\lim_{n\to\infty} k_n^{6\alpha-3} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(2)}(y) e^{-\alpha y}\, \mathrm{d}x\, \mathrm{d}y = 0.$$

Similarly, $2\alpha - 1 > 1/2$ for $\alpha > 3/4$ and we get

$$k_n \cdot n^{1-2\alpha+\varepsilon} \ll n^{-1/2+\varepsilon} \cdot k_n = o(1),$$

provided that ε is small enough, so that

$$\lim_{n\to\infty} k_n^{2\alpha} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(2)}(y) e^{-\alpha y} \,\mathrm{d}x\,\mathrm{d}y = 0.$$

For the third sub-domain $\mathcal{D}_3$ we use (4.106) which states that if $p_2 = (x_2, y_2) \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1) \cap \mathcal{B}(y)$ and $y_2 \le R - y_1$, then $y_2 \ge \hat{y}(x_1, y_1)$, where $\hat{y}(x_1, y_1) = \big(2\log(|x_1| e^{-y_1/2}) - \log c_K\big) \vee 0$. Moreover, $p_2 \in \check{\mathcal{B}}_{\mathrm{Po}}(p_1)$.

Again, we will use the Campbell-Mecke formula to obtain

$$\begin{aligned}
&\mathcal{I}_n^{(3)}(y)\\
&= \mathbb{E}\left[\sum_{(p_1,p_2)\in\mathcal{D}_3} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}} \cdot \mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}} \cdot D_{\mathrm{Po}}(y, k_n-2; \mathcal{P}\setminus\{p_1,p_2\})\right]\\
&= O(1)\rho_{\mathrm{Po}}(y, k_n-2) \int_{-I_n}^{I_n} \int_{(1-\varepsilon)R\wedge(R-y)}^{R} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}} \times\\
&\qquad \int_{-I_n}^{I_n} \int_0^{R-y_1} \mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}} e^{-\alpha(y_1+y_2)} dy_2 dx_2 dy_1 dx_1.
\end{aligned}$$

The inner integral with respect to $p_2 := (x_2, y_2)$ is

$$\begin{aligned}
&\int_{-I_n}^{I_n} \int_0^{R-y_1} \mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}} e^{-\alpha y_2} dy_2 dx_2\\
&\qquad \le \int_{-I_n}^{I_n} \int_0^{R-y_1} \mathbb{1}_{\left\{y_2\ge\hat{y}(x_1,y_1),\, p_2\in\check{\mathcal{B}}_{\mathrm{Po}}((0,y))\right\}} e^{-\alpha y_2} dy_2 dx_2\\
&\qquad = O(1) e^{y/2} \int_{\hat{y}(x_1,y_1)}^{R-y_1} e^{y_2/2-\alpha y_2} dy_2\\
&\qquad = O(1) e^{y/2+(1/2-\alpha)\hat{y}(x_1,y_1)}.
\end{aligned}$$

Thus, we get

$$\begin{aligned}
&\int_{-I_n}^{I_n} \int_{(1-\varepsilon)R\wedge(R-y)}^{R} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\}} \int_{-I_n}^{I_n} \int_0^{R-y_1} \mathbb{1}_{\{p_2\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p_1)\cap\mathcal{B}(y)\}} \times\\
&\qquad\qquad e^{-\alpha(y_1+y_2)} dy_2 dx_2 dy_1 dx_1\\
&\le O(1) \int_{-I_n}^{I_n} \int_{(1-\varepsilon)R\wedge(R-y)}^{R} e^{y/2+(1/2-\alpha)\hat{y}(x_1,y_1)} e^{-\alpha y_1} dy_1 dx_1.
\end{aligned}$$

Due to symmetry, to bound the integral it is enough to integrate this with respect to x_1 from 0 to I_n. We will split this integral into two parts according to the value of $c(x_1, y_1)$ as

$$\int_0^{I_n} e^{\hat{y}(x_1,y_1)(1/2-\alpha)} dx_1 = \int_{c_K e^{y_1/2}}^{I_n} e^{c(x_1,y_1)(1/2-\alpha)} dx_1 + \int_0^{c_K e^{y_1/2}} dx_1.$$

The first integral becomes

$$\int_{c_K e^{y_1/2}}^{I_n} e^{\hat{y}(x_1,y_1)(1/2-\alpha)} dx_1 = O(1) \cdot \int_{c_K e^{y_1/2}}^{I_n} x_1^{1-2\alpha} e^{-\frac{y_1}{2}(1-2\alpha)} dx_1$$
$$= \begin{cases} O(R) \cdot e^{-y_1/2+\alpha y_1} \cdot e^{\frac{R}{2}2(1-\alpha)} & \text{if } \alpha \le 1, \\ O(1) \cdot e^{-y_1/2+\alpha y_1 + 2(1-\alpha)y_1/2} & \text{if } \alpha > 1, \end{cases}$$
$$= \begin{cases} O(R) \cdot e^{(\alpha-1/2)y_1} \cdot n^{2(1-\alpha)} & \text{if } \alpha \le 1, \\ O(1) \cdot e^{y_1/2} & \text{if } \alpha > 1. \end{cases}$$

The second integral trivially gives

$$\int_0^{c_K e^{y_1/2}} dx_1 = O(1) \cdot e^{y_1/2}.$$

Putting these two together we conclude that

$$e^{y/2} \cdot \int_0^{I_n} e^{\hat{y}(x_1,y_1)(1/2-\alpha)} dx_1 = O(1) \cdot e^{y_1/2 + y(3/2-\alpha)}.$$

Now, we integrate these with respect to y_1 as

$$n^{2(1-\alpha)} \cdot \int_{(1-\varepsilon)R \wedge (R-y)}^{R} e^{(\alpha-1/2)y_1 - \alpha y_1} dy_1 = O(1) \cdot n^{2(1-\alpha)} \cdot e^{-R/2+\varepsilon R/2 + y/2}$$
$$= O(1) \cdot n^{1-2\alpha+\varepsilon} \cdot e^{y/2}.$$

Therefore, we conclude that

$$\mathcal{I}_n^{(3)}(y) = O\left(R\right) n^{1-2\alpha+\varepsilon(2\alpha-1)}\, e^{y/2} \rho_{\mathrm{Po}}(y, k_n - 2),$$

and hence, again using Lemma 4.7.4,

$$\int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(3)}(y) e^{-\alpha y} \, \mathrm{d}x \, \mathrm{d}y$$
$$= O\left(R\right) n^{1-2\alpha+\varepsilon(2\alpha-1)} \int_{\mathcal{K}_C(k_n)} e^{y/2} \rho_{\mathrm{Po}}(y, k_n - 2) e^{-\alpha y} \, \mathrm{d}x \, \mathrm{d}y$$
$$= O\left(R\right) n^{1-2\alpha+\varepsilon(2\alpha-1)} k_n^{-2\alpha+1}.$$

It follows that for $\varepsilon = \varepsilon(\alpha)$ small enough

$$k_n^{6\alpha-3} R n^{1-2\alpha+\varepsilon(2\alpha-1)} k_n^{-2\alpha+1} = R n^{\varepsilon(2\alpha-1)} \left(\frac{k_n^2}{n}\right)^{2\alpha-1} = o\left(1\right),$$

and hence, for $\alpha > 1/2$,

$$\lim_{n\to\infty} k_n^{6\alpha-3} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(3)}(y) e^{-\alpha y} \, \mathrm{d}x \, \mathrm{d}y = 0.$$

Since $2\alpha - 1 \geq 1/2$ when $\alpha \geq 3/4$ it immediately follows that

$$\lim_{n\to\infty} k_n^{2\alpha} \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(3)}(y) e^{-\alpha y} \, \mathrm{d}x \, \mathrm{d}y = 0.$$

The sums (4.91) **and** (4.92) Again, the analysis for both terms are similar and we shall analyze (4.91) only. Let us set $p = (0, y)$. Recall that $\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y) \cap \mathcal{R}([R - y + 2\log\left(\frac{\pi}{2}\right), R]) = \emptyset$. Thus, the summand in (4.91) is equal to 0, when $y_1 > R - y + 2\log(\pi/2)$.

Recall the definition of the extended ball $\check{\mathcal{B}}_{\mathrm{Po}}(p)$ around p in (4.104) that contains both $\mathcal{B}(p)$ and $\mathcal{B}_\infty(p)$, i.e.,

$$\check{\mathcal{B}}_{\mathrm{Po}}(y) := \{p' : y' < R - y, \; |x'| < (1+K)e^{\frac{1}{2}(y+y')}\},$$

and that we have $\mathbb{E}\left[D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\})\right] = \rho_{\mathrm{Po}}(y, k_n - 2)$.

Further, observe that,

$$\mathcal{B}(y) \cap \mathcal{R}([0, R - y)) \subseteq \check{\mathcal{B}}_{\mathrm{Po}}(y),$$

and

$$\mathcal{B}(y) \cap \mathcal{R}([R - y, R]) = \mathcal{R}([R - y, R]).$$

We thus conclude that

$$\mathcal{B}(y) \subseteq \check{\mathcal{B}}_{\mathrm{Po}}(y) \cup \mathcal{R}([R - y, R]). \tag{4.108}$$

Hence, if we set

$$h_y(p_1) := \mathbb{1}_{\{p_1 \in \mathcal{B}(p) \setminus \mathcal{B}_\infty(y)\}} \cdot \left(\mu\left(\check{\mathcal{B}}_{\mathrm{Po}}(p_1) \cap \check{\mathcal{B}}_{\mathrm{Po}}(y)\right) + \mu\left(\mathcal{R}([R - y, R])\right)\right),$$

then

$$\begin{aligned}
&\mathbb{1}_{\{p_1 \in \mathcal{B}(p) \setminus \mathcal{B}_\infty(y)\}} \cdot \mathbb{E}\left[\left(\sum_{p_2 \in \mathcal{P} \setminus \{p, p_1\}} \mathbb{1}_{\{p_2 \in \mathcal{B}(y) \cap \mathcal{B}_\infty(p_1)\}}\right) \cdot D_{\mathrm{Po}}(y, k_n - 2; \mathcal{P} \setminus \{p_1, p_2\})\right] \\
&= O(1) \cdot \mathbb{1}_{\{p_1 \in \mathcal{B}(y) \setminus \mathcal{B}_\infty(y)\}} \cdot \mu(\mathcal{B}(y) \cap \mathcal{B}(p_1)) \rho_{\mathrm{Po}}(y, k_n - 2) \\
&\leq O(1) \cdot h_y(p_1) \rho_{\mathrm{Po}}(y, k_n - 2).
\end{aligned}$$

To calculate the expectation of the above function we need to approximate the intersection of the two balls $\check{\mathcal{B}}_{\mathrm{Po}}(y)$ and $\check{\mathcal{B}}_{\mathrm{Po}}(p_1)$, where $p_1 = (x_1, y_1)$. Let us assume without loss of generality that $x_1 > 0$. The right boundary of $\check{\mathcal{B}}_{\mathrm{Po}}(y)$ is given by the equation $x = x(y') = (1+K)e^{\frac{1}{2}(y+y')}$ whereas the left boundary of $\check{\mathcal{B}}_{\mathrm{Po}}(p_1)$ is given by the curve $x = x(y') = x_1 - (1+K)e^{\frac{1}{2}(y_1+y')}$.

The equation that determines the intersection point of the two curves is

$$x_1 - (1+K)e^{(\hat{y}+y_1)/2} = (1+K)e^{(\hat{y}+y)/2},$$

where $\hat{y}$ is the y-coordinate of the intersection point. We can solve the above for $\hat{y}$ as

$$x_1 = (1+K)e^{\hat{y}/2}\left(e^{y/2} + e^{y_1/2}\right).$$

But since $p_1 = (x_1, y_1) \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p)$, we also have $x_1 > e^{\frac{y+y_1}{2}}$. Therefore,

$$e^{\hat{y}/2} > \frac{1}{1+K}\,\frac{e^{\frac{y+y_1}{2}}}{e^{y/2}+e^{y_1/2}} \geq \frac{1}{2(1+K)}\,\frac{e^{\frac{y_1+y}{2}}}{e^{(y\vee y_1)/2}} > \frac{1}{2(1+K)}\,e^{(y\wedge y_1)/2}. \tag{4.109}$$

The above yields

$$\hat{y} > (y\wedge y_1) - 2\log(2(1+K)) := \hat{y}(y_1, y), \tag{4.110}$$

which, in turn, implies the following

$$p \in \check{\mathcal{B}}_{\mathrm{Po}}((0,y)) \cap \check{\mathcal{B}}_{\mathrm{Po}}(p_1) \Rightarrow y(p) \geq \hat{y}(y_1, y). \tag{4.111}$$

We thus conclude that

$$\mathcal{B}\,(p_1) \cap \mathcal{B}\,(p) \subseteq \left(\check{\mathcal{B}}_{\mathrm{Po}}(p) \cap \mathcal{R}([\hat{y}(y_1,y), R])\right) \cup \mathcal{R}([R-y, R]),$$

which in turn implies that

$$\mu\left(\check{\mathcal{B}}_{\mathrm{Po}}(p_1) \cap \mathcal{B}\,(p)\right) \leq \mu\left(\check{\mathcal{B}}_{\mathrm{Po}}(p) \cap \mathcal{R}([\hat{y}(y_1,y), R]\right) + \mu(\mathcal{R}([R-y, R])).$$

Therefore,

$$\begin{aligned} h_y(p_1, \mathcal{P}) \leq\ & \mathbb{1}_{\{p_1 \in \mathcal{B}(p)\setminus\mathcal{B}_\infty(p)\}}\mu\left(\check{\mathcal{B}}_{\mathrm{Po}}(p) \cap \mathcal{R}([\hat{y}(y_1,y), R])\right) \\ & + \mathbb{1}_{\{p_1 \in \mathcal{B}(p)\setminus\mathcal{B}_\infty(p)\}}\mu\left(\mathcal{R}([R-y, R])\right). \end{aligned}$$

Now, the Campbell-Mecke formula gives

$$\begin{aligned} & \mathbb{E}\left[\sum_{\substack{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}\\ y(p_1)\geq K}} \mathbb{1}_{\{p_1\in\mathcal{B}(y)\setminus\mathcal{B}_\infty(y)\}}\mathbb{1}_{\{p_2\in\mathcal{B}(y)\cap\mathcal{B}_\infty(y)\}}\, D_{\mathrm{Po}}(y, k_n-2; \mathcal{P}\setminus\{p_1,p_2\})\right] \\ & \leq \mathbb{E}\left[\left(\sum_{p_1\in\mathcal{P}} h_y(p_1, \mathcal{P}\setminus\{p_1\})\right)\right] \\ & = \frac{\nu\alpha}{\pi}\int_{\mathcal{R}} \mathbb{E}\left[h_y(p_1, \mathcal{P}\setminus\{p_1\})\right] e^{-\alpha y_1}\,\mathrm{d}x_1\,\mathrm{d}y_1 \\ & \leq \frac{\nu\alpha}{\pi}\int_{\mathcal{R}} \mathbb{1}_{\{p_1\in\mathcal{B}(p)\setminus\mathcal{B}_\infty(p)\}}\mu\left(\check{\mathcal{B}}_{\mathrm{Po}}(p)\cap\mathcal{R}([\hat{y}(y_1,y),R])\right) e^{-\alpha y_1}\,\mathrm{d}x_1\,\mathrm{d}y_1 \qquad (4.112) \\ & \quad + \frac{\nu\alpha}{\pi}\int_{\mathcal{R}} \mathbb{1}_{\{p_1\in\mathcal{B}(p)\setminus\mathcal{B}_\infty(p)\}}\mu\left(\mathcal{R}([R-y,R])\right) e^{-\alpha y_1}\,\mathrm{d}x_1\,\mathrm{d}y_1. \qquad (4.113) \end{aligned}$$

Recall that $(\mathcal{B}_{\mathrm{Po}\,\triangle\infty}((0,y))) \cap \mathcal{R}([R-y+2\log\left(\frac{\pi}{2}\right), R]) = \emptyset$. We will first calculate the measures μ appearing in (4.112) and (4.113). The first one is

$$\mu\left(\check{\mathcal{B}}_{\mathrm{Po}}(y) \cap \mathcal{R}([c(y_1,y),R])\right) \leq (1+K)\frac{\nu\alpha}{\pi}\cdot e^{y/2}\int_{\hat{y}(y_1,y)}^{R} e^{-(\alpha-\frac{1}{2})y'}\,dy'$$
$$= O\left(e^{\frac{y}{2}-(\alpha-\frac{1}{2})(y\wedge y_1)}\right).$$

The second term is

$$\mu\left(\mathcal{R}([R-y,R])\right) = \frac{\nu\alpha}{\pi}\int_{R-y}^{R} \pi e^{\frac{R}{2}} e^{-\alpha y'}\,dy' = O\left(e^{\frac{R}{2}}e^{-\alpha(R-y)}\right) = O\left(e^{\alpha y-(\alpha-\frac{1}{2})R}\right).$$

Using these, we get

$$\int_{\mathcal{R}([0,R-y_n+2\ln\frac{\pi}{2}])} \mathbb{E}\left[h_y(p_1,\mathcal{P}\setminus\{p_1\})\right] e^{-\alpha y_1}\,dx_1\,dy_1$$
$$= O(1)\int_{\mathcal{R}([0,R-y+2\ln\frac{\pi}{2}])} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(p)\}} e^{\frac{y}{2}-(\alpha-\frac{1}{2})(y\wedge y_1)-\alpha y_1}\,dx_1\,dy_1 \qquad (4.114)$$
$$+\,O(1)\int_{\mathcal{R}([0,R-y+2\ln\frac{\pi}{2}])} \mathbb{1}_{\{p_1\in\mathcal{B}((0,y))\}} e^{\alpha y-(\alpha-\frac{1}{2})R-\alpha y_1}\,dx_1\,dy_1. \qquad (4.115)$$

Now, Lemma 1.6.2 implies that for any $y_1 \in [0, R-y+2\ln\frac{\pi}{2}]$, we have

$$\int_{-I_n}^{I_n} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)\}}\,dx_1 \leq 2Ke^{\frac{3}{2}(y_1+y)-R}.$$

Therefore, (4.114) is

$$O(1)\cdot e^{2y-R}\int_0^{R-y+2\ln\frac{\pi}{2}} e^{\frac{3y_1}{2}-(\alpha-\frac{1}{2})(y_1\wedge y)-\alpha y_1}\,dy_1$$
$$= O(1)\cdot e^{2y-R}\left(\int_0^{y} e^{\frac{3y_1}{2}-(2\alpha-\frac{1}{2})y_1}\,dy_1 + e^{-(\alpha-\frac{1}{2})y}\int_y^{R-y+2\ln\frac{\pi}{2}} e^{(\frac{3}{2}-\alpha)y_1}\,dy_1\right)$$
$$= O(1)\left(\begin{cases} e^{(4-2\alpha)y-R}, & \text{if } \alpha<1 \\ R\cdot e^{2y-R}, & \text{if } \alpha\geq 1\end{cases} + \begin{cases} e^{-(\alpha-\frac{1}{2})R+y}, & \text{if } \alpha<3/2 \\ R\cdot e^{2(2-\alpha)y-R}, & \text{if } \alpha\geq 3/2\end{cases}\right).$$

Similarly, for (4.115), we have

$$\int_{\mathcal{R}([0,R-y+2\ln\frac{\pi}{2}])} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}((0,y))\}} e^{\alpha y-(\alpha-\frac{1}{2})R-\alpha y_1}\,dx_1\,dy_1$$
$$= e^{\frac{3y}{2}-R+\alpha y-(\alpha-\frac{1}{2})R}\cdot\int_0^{R-y+2\ln\frac{\pi}{2}} e^{\frac{3y_1}{2}-\alpha y_1}\,dy_1$$

$$= O(1) \cdot \begin{cases} e^{\frac{3y}{2} - R + \alpha y - (\alpha - \frac{1}{2})R + (\frac{3}{2} - \alpha)(R-y)}, & \text{if } \alpha < 3/2, \\ R \cdot e^{(\frac{3}{2}+\alpha)y - (\alpha + \frac{1}{2})R}, & \text{if } \alpha \geq 3/2, \end{cases}$$

$$= O(1) \cdot \begin{cases} e^{-(2\alpha-1)R + 2\alpha y}, & \text{if } \alpha < 3/2, \\ R \cdot e^{(\frac{3}{2}+\alpha)y - (\alpha + \frac{1}{2})R}, & \text{if } \alpha \geq 3/2. \end{cases}$$

We thus conclude, using $2(2-\alpha)y \leq y$ for $\alpha > 3/2$, that

$$\mathbb{E}\left[\left(\sum_{p_1 \in \mathcal{P}\setminus\{p\}} h_y(p_1, \mathcal{P}\setminus\{p_1\})\right)\right] \leq O(1) \cdot \left(\mathcal{I}_n^{(1)}(y) + \mathcal{I}_n^{(2)}(y) + \mathcal{I}_n^{(3)}(y)\right), \quad (4.116)$$

where

$$\mathcal{I}_n^{(1)}(y) = \begin{cases} e^{(4-2\alpha)y - R}, & \text{if } \alpha < 1, \\ R \cdot e^{2y - R}, & \text{if } \alpha \geq 1, \end{cases}$$

$$\mathcal{I}_n^{(2)}(y) = \begin{cases} e^{-(\alpha - \frac{1}{2})R + y}, & \text{if } \alpha < 3/2, \\ R \cdot e^{y - R}, & \text{if } \alpha \geq 3/2, \end{cases}$$

$$\mathcal{I}_n^{(3)}(y) = \begin{cases} e^{-(2\alpha-1)R + 2\alpha y}, & \text{if } \alpha < 3/2, \\ R \cdot e^{(\frac{3}{2}+\alpha)y - (\alpha + \frac{1}{2})R}, & \text{if } \alpha \geq 3/2. \end{cases}$$

We proceed to calculate

$$\int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[\left(\sum_{p_1 \in \mathcal{P}} h_y(p_1, \mathcal{P} \setminus \{p_1\})\right)\right] \cdot \rho_{\mathrm{Po}}(y, k_n - 2) e^{-\alpha y} \, \mathrm{d}y.$$

For this we define

$$M_i = \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(i)}(y) \rho_{\mathrm{Po}}(y, k_n - 1) e^{-\alpha y} \, \mathrm{d}y,$$

so that

$$\int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[\left(\sum_{p_1 \in \mathcal{P}\setminus\{(0,y)\}} h_y(p_1, \mathcal{P}\setminus\{p_1\})\right)\right] \rho_{\mathrm{Po}}(y, k_n - 1) e^{-\alpha y} \, \mathrm{d}y$$
$$= O\left(M_1 + M_2 + M_3\right).$$

Computing each of the integrals separately we obtain, using Lemma 4.7.4 and the fact that $n = \nu e^{R/2}$,

$$M_1 := \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(1)}(y) \rho_{\mathrm{Po}}(y, k_n - 1) e^{-\alpha y} \, \mathrm{d}y = O(1) \cdot \begin{cases} \frac{k_n^{7-6\alpha}}{n^2}, & \text{if } \alpha < 1, \\ R \frac{k_n^{3-2\alpha}}{n^2}, & \text{if } \alpha \geq 1, \end{cases}$$

$$M_2 := \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(2)}(y)\rho_{\mathrm{Po}}(y, k_n - 1)e^{-\alpha y}dy = O(1) \cdot \begin{cases} \frac{k_n^{1-2\alpha}}{n^{2\alpha-1}}, & \text{if } \alpha < 3/2, \\ R\frac{k_n^{1-2\alpha}}{n^2}, & \text{if } \alpha \geq 3/2, \end{cases}$$

and finally

$$M_3 := \int_{\mathcal{K}_C(k_n)} \mathcal{I}_n^{(3)}(y)\rho_{\mathrm{Po}}(y, k_n - 1)e^{-\alpha y}dy = O(1) \cdot \begin{cases} \frac{k_n^{2\alpha-1}}{n^{4\alpha-2}}, & \text{if } \alpha < 3/2, \\ R \cdot \frac{k_n^2}{n^{2\alpha+1}}, & \text{if } \alpha \geq 3/2. \end{cases}$$

Now, we will consider the two cases according to the value of α. First we note that $R = O(\log(n))$ and since $k_n = O(n^{\frac{1}{2\alpha+1}})$ and $\alpha > 1/2$ we have that $Rk_n^2 n^{-1} = o(1)$. Assume first that $1/2 < \alpha < 3/4$. In this case, we want to show that

$$\lim_{n\to\infty} k_n^{6\alpha-3}(M_1 + M_2 + M_3) = 0. \tag{4.117}$$

Using the above expression for M_i, we have

$$k_n^{6\alpha-3}(M_1 + M_2 + M_3) = O(1) \cdot k_n^{6\alpha-3}\left(\frac{k_n^{7-6\alpha}}{n^2} + \frac{k_n^{1-2\alpha}}{n^{2\alpha-1}} + \frac{k_n^{2\alpha-1}}{n^{4\alpha-3}}.\right)$$

We wish to show that each one of the above three terms is $o(1)$ for $k_n = O(n^{\frac{1}{2\alpha+1}})$. For the first one we have

$$k_n^{6\alpha-3}\frac{k_n^{7-6\alpha}}{n^2} = \left(\frac{k_n^2}{n}\right)^2 = o(1).$$

The second term yields

$$k_n^{6\alpha-3}\frac{k_n^{-2\alpha+1}}{n^{2\alpha-1}} = \left(\frac{k_n^2}{n}\right)^{2\alpha-1} = o(1).$$

Finally, the third one yields

$$k_n^{6\alpha-3} \cdot \frac{k_n^{2\alpha-1}}{n^{4\alpha-2}} = \left(\frac{k_n^2}{n}\right)^{4\alpha-2} = o(1).$$

For $\alpha \geq 3/4$, we would like to show that

$$\lim_{n\to\infty} k_n^{2\alpha} \cdot (M_1 + M_2 + M_3) = 0. \tag{4.118}$$

Firstly, we note that each M_i is as above if $3/4 < \alpha < 1$. Therefore, since for this range $2\alpha < 6\alpha - 3$ the result follows from the above analysis. Next we consider the case $1 \leq \alpha < 3/2$. Here, only the value of M_1 changes and we compute that

$$k_n^{6\alpha-3}M_1 = O(1)\log(n)n^{-2}k_n^{4\alpha} \leq O(\log(n))\left(\frac{k_n^2}{n}\right)^2 = o(1),$$

so that (4.118) holds for $3/4 < \alpha < 1$.

Proceeding with the case $\alpha \geq 3/2$, it is only M_2 and M_3 that change values. In particular, for any $\alpha \geq 3/2$ we have

$$\frac{k_n}{n} M_2 = O(1) R \frac{k_n}{n^2} = o(1).$$

Also,

$$k_n^{2\alpha} M_3 = O(1) R \frac{k_n^{2\alpha+2}}{n^{2\alpha+1}} = Ro\left(\frac{n^{\alpha+1}}{n^{2\alpha+1}}\right) = o(1),$$

since $k_n = o(n^{1/2})$ and hence (4.118) holds. This finished the proof for (4.91).

The sum of (4.93) Using the Campbell-Mecke formula, we write

$$\begin{aligned}
&\mathbb{E}\left[\sum_{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\},\ y_1<K}^{\neq} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)\}}\mathbb{1}_{\{p_2\in\mathcal{B}(y)\cap\mathcal{B}_\infty(y)\}}\right] \\
&\leq \int_0^K \int_{-I_n}^{I_n} \int_0^R \int_{-I_n}^{I_n} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)\}}\mathbb{1}_{\{p_2\in\mathcal{B}(y)\cap\mathcal{B}_\infty(y)\}} e^{-\alpha y_2} e^{-\alpha y_1}\, \mathrm{d}x_2\, \mathrm{d}y_2\, \mathrm{d}x_1\, \mathrm{d}y_1 \\
&\leq \mu(\mathcal{B}(y)) \cdot \int_{-I_n}^{I_n} \int_0^K \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)\}} e^{-\alpha y_1} dx_1 dy_1.
\end{aligned}$$

Recall that by Lemma 3.3.4, $\mu(\mathcal{B}(y)) = O(1)e^{y/2}$. We bound the integral using Lemma 1.6.2. In particular, (1.9) implies that if $p_1 = (x_1, y_1) \in \mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)$, then because $y_1 < K$

$$|x_1 - e^{(y+y_1)/2}| \leq e^{(y+y_1)/2} \cdot K e^{y+y_1-R} = O(1) e^{(y+y_1)/2} \cdot e^{y-R}.$$

Therefore,

$$\begin{aligned}
&\int_{-I_n}^{I_n} \int_0^K \mathbb{1}_{\{(x_1,y_1)\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}((0,y))\}} e^{-\alpha y_1} dx_1 dy_1 \\
&= O(1) \cdot e^{y-R} \cdot \int_0^K e^{(y+y_1)/2} e^{-\alpha y_1} dy_1 \\
&= O(1) \cdot e^{3y/2-R},
\end{aligned}$$

and hence

$$\mathbb{E}\left[\sum_{\substack{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}\\ y_1<K}}^{\neq} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)\}}\mathbb{1}_{\{p_2\in\mathcal{B}(y)\cap\mathcal{B}_\infty(y)\}}\right] = O(1) \cdot e^{2y-R}.$$

Now, we integrate this over y to obtain that

$$\int_{\mathcal{K}_C(k_n)} \mathbb{E}\left[\sum_{\substack{p_1,p_2\in\mathcal{P}\setminus\{(0,y)\}\\ y_1<K}}^{\neq} \mathbb{1}_{\{p_1\in\mathcal{B}_{\mathrm{Po}\,\triangle\infty}(y)\}}\mathbb{1}_{\{p_2\in\mathcal{B}(y)\cap\mathcal{B}_\infty(y)\}}\right] e^{-\alpha y}\,\mathrm{d}y$$
$$= O(1)e^{-R}\int_{\mathcal{K}_C(k_n)} e^{2y-\alpha y}\,\mathrm{d}y = O(1)n^{-2}\begin{cases} k_n^{4-2\alpha}, & \text{if } \alpha<2,\\ \log k_n, & \text{if } \alpha=2,\\ 1, & \text{if } \alpha>2.\end{cases}$$

To finish the argument assume first that $1/2<\alpha\leq 3/4$. In this case,

$$k_n^{6\alpha-3}n^{-2}k_n^{4-2\alpha} = n^{-2}\cdot k_n^{4\alpha+1} = o(1).$$

For $3/4\leq\alpha<2$ we use that $2\alpha<6\alpha-3$, so that $k_n^{2\alpha}n^{-2}k_n^{4-2\alpha}=o(1)$. Finally, when $\alpha\geq 2$, we have that

$$k_n^{2\alpha}(\log(k_n)\wedge 1)n^{-2}\leq k_n^{2\alpha+1}n^{-2} = O\left(n^{-1}\right) = o(1).$$

which completes the proof for (4.93) and thus the proof of Proposition 4.3.4. □

4.7.3 Coupling G_n to G_{Po}

Now that we have established the equivalence of the clustering function between the Poissonized KPKVB graph G_{Po} and the finite box graph G_{box} the final step is to relate the clustering function in G_{Po} to the KPKVB graph G_n. As mentioned in Section 4.3.2, this is done by moving from $c(k_n;G_n)$ to the adjusted clustering function $c^*(k_n;G_n)$ (Lemma 4.3.2) and then to $c^*(k_n;G_{\mathrm{Po}})$ (Proposition 4.3.3). We start with a technical lemma on the difference between the number of vertices with degree k_n in both models. Then we give the proof of Proposition 4.3.3 and after that we prove Lemma 4.3.2.

Lemma 4.7.5. *Let $(k_n)_{n\geq1}$ be sequence of natural numbers with $k_n=o(n^{\frac{1}{2\alpha+1}})$. Then, on the coupling space described in Section 1.6.4,*

$$\mathbb{E}\left[|N_n(k_n)-N_{\mathrm{Po}}(k_n)|\right] = o\left(\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]\right) = o\left(nk^{-(2\alpha+1)}\right),$$

and in particular,

$$\mathbb{E}\left[N_n(k_n)\right] = \Theta\left(nk_n^{-(2\alpha+1)}\right).$$

Proof. The second claim follows immediately from the first one and the fact that $\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]=\Theta\left(nk_n^{-(2\alpha+1)}\right)$, see Lemma 4.7.3.

Denote by $\mathcal{V}_n(k_n)$ and $\mathcal{V}_{\mathrm{Po}}(k_n)$ the set of points u_i with degree k_n in G_n and G_{Po}, respectively. Then the result follows if we can show that

$$\mathbb{E}\left[\sum_{i=1}^{n\wedge N}\mathbb{1}_{\{u_i\in\mathcal{V}_n(k_n)\Delta\mathcal{V}_{\mathrm{Po}}(k_n)\}}\middle|\, N\right] = o(n)\,\mathbb{P}\left(\deg_{N-1}(U)=k_n\middle|\, N\right)$$

We first use a Chernoff-based large deviation result for a Poisson random variable (1.12), which implies that for any $C > 0$ we have $N \in [n - C\sqrt{n \log n}, n + C\sqrt{n \log n}]$ with probability $1 - n^{-C^2/2}$. Since we can select $C > 0$ arbitrarily large it follows that we only need to prove the result conditionally on the event $|N - n| \le C\sqrt{n \log(n)}$. Note that if $N = n$ then

$$\sum_{i=1}^{n \wedge N} \mathbb{1}_{\{u_i \in \mathcal{V}_n(k_n) \Delta \mathcal{V}_{\mathrm{Po}}(k_n)\}} = 0.$$

Hence if we denote by A_n^- the event that $n - C\sqrt{n \log(n)} \le N < n$ and by A_n^+ the event that $n < N \le n + C\sqrt{n \log(n)}$, we need to show that

$$\mathbb{E}\left[\sum_{i=1}^{n \wedge N} \mathbb{1}_{\{u_i \in \mathcal{V}_n(k_n) \Delta \mathcal{V}_{\mathrm{Po}}(k_n)\}} \middle| A_n^+\right] = o(n)\, \mathbb{P}\left(\deg_N(U) = k_n \middle| A_n^+\right),$$

and likewise with A_n^+ replaced by A_n^-.

We first consider the case where $n < N \le n + C\sqrt{n \log(n)}$, i.e. the case where we have more vertices in G_{Po} than in G_n. The proof for the other case is similar and we omit it. Let $W = \{u_{n+1}, \dots, u_N\}$ be the set of vertices in G_{Po} that are not in G_n. To ease notation we let $\mathbb{E}_n$ and $\mathbb{P}_n$ denote the conditional expectation and probability on the event A_n^+. First we note that

$$\begin{aligned}
\mathbb{E}_n\left[\sum_{i=n+1}^{N} \mathbb{1}_{\{u_i \in \mathcal{V}_n(k_n) \Delta \mathcal{V}_{\mathrm{Po}}(k_n)\}}\right] &= \mathbb{E}_n\left[\sum_{i=n+1}^{N} \mathbb{1}_{\{\deg_{Po}(u_i) = k_n\}}\right] \\
&= \frac{(N-n)}{n} \mathbb{1}_{A_n^+} n \mathbb{P}_n\left(\deg_N(U) = k_n\right) \\
&\le C\sqrt{\frac{\log(n)}{n}} n \mathbb{P}_n\left(\deg_N(U) = k_n\right) \\
&= o(n)\, \mathbb{P}_n\left(\deg_N(U) = k_n\right).
\end{aligned}$$

Hence we are left to consider the sum over all vertices in $\mathcal{V}_n$.

Let $u_i \in \mathcal{V}_n(k_n)$. Then since $N > n$ we have that $u_i \in \mathcal{V}_{\mathrm{Po}}(k_n)$ and hence $u_i \in \mathcal{V}_n(k_n) \Delta \mathcal{V}_{\mathrm{Po}}(k_n)$ if and only if $|B_{\mathbb{H}}(u_i) \cap W| \ge 1$ and either $u_i \in \mathcal{V}_n(k_n)$ or $u_i \in \mathcal{V}_{\mathrm{Po}}(k_n)$. For any $u \in \mathbb{H}$ let $Q_n(u)$ denote the event that $|\mu_{\mathbb{H}}(u) - k_n| \le C\sqrt{k_n \log(k_n)}$ and denote by $Q_n(u)^c$ its complement. Then,

$$\begin{aligned}
&\sum_{i=1}^{n} \left(\mathbb{1}_{\{\deg_n(u_i) = k_n\}} + \mathbb{1}_{\{\deg_{\mathrm{Po}}(u_i) = k_n\}}\right) \mathbb{1}_{\{|B_{\mathbb{H}}(u_i) \cap W| \ge 1\}} \\
&\le \sum_{i=1}^{n} \left(\mathbb{1}_{\{\deg_n(u_i) = k_n\}} + \mathbb{1}_{\{\deg_{\mathrm{Po}}(u_i) = k_n\}}\right) \mathbb{1}_{\{|B_{\mathbb{H}}(u_i) \cap W| \ge 1\}} \mathbb{1}_{Q_n(u_i)} \\
&\quad + \sum_{i=1}^{n} \left(\mathbb{1}_{\{\deg_n(u_i) = k_n\}} + \mathbb{1}_{\{\deg_{\mathrm{Po}}(u_i) = k_n\}}\right) \mathbb{1}_{Q_n(u_i)^c}.
\end{aligned}$$

The second sum splits into two parts. The expectation of the first part yields

$$\begin{aligned}\mathbb{E}_n\left[\sum_{i=1}^n \mathbb{1}_{\{\deg_n(u_i)=k_n\}}\mathbb{1}_{\{Q_n(u_i)^c\}}\right] &= n\,\mathbb{P}\left(\mathrm{Bin}\left(n-1,\frac{\mu_{\mathbb{H}}(U)}{n}\right)=k_n, Q_n(U)^c\right)\\ &= O\left(nk_n^{-\frac{C^2}{3}}\right)=o\left(nk_n^{-(2\alpha+1)}\right),\end{aligned}$$

where we have used Lemma D.1 in the appendix, for any $C^2/3 > 2\alpha+1$. Similarly, Lemma D.1 implies that the expectation of the second part is

$$\begin{aligned}\mathbb{E}_n\left[\sum_{i=1}^n \mathbb{1}_{\{\deg_{\mathrm{Po}}(u_i)=k_n\}}\mathbb{1}_{\{Q_n(u_i)^c\}}\right] &= n\,\mathbb{P}\left(\mathrm{Bin}\left(N-1,\frac{\mu_{\mathbb{H}}(U)}{n}\right)=k_n, Q_n(U)^c\right)\\ &= O\left(nk_n^{-\frac{C^2}{3}}\right)=o\left(nk_n^{-(2\alpha+1)}\right).\end{aligned}$$

Hence it remains to show that

$$\mathbb{E}_n\left[\sum_{i=1}^n\left(\mathbb{1}_{\{\deg_n(u_i)=k_n\}}+\mathbb{1}_{\{\deg_{\mathrm{Po}}(u_i)=k_n\}}\right)\mathbb{1}_{\{|B_{\mathbb{H}}(u_i)\cap W|\geq 1\}}\mathbb{1}_{Q_n(u_i)}\right]=o\left(nk_n^{-(2\alpha+1)}\right).$$

We will show that

$$\mathbb{E}_n\left[\sum_{i=1}^n \mathbb{1}_{\{\deg_n(u_i)=k_n\}}\mathbb{1}_{\{|B_{\mathbb{H}}(u_i)\cap W|\geq 1\}}\mathbb{1}_{Q_n(u_i)}\right]=o\left(nk_n^{-(2\alpha+1)}\right).$$

The other term follows using almost identical arguments and is omitted. To prove the above result we note that since $N > n$ we have $\mathbb{1}_{\{|B_{\mathbb{H}}(u_i)\cap W|\geq 1\}} \leq \sum_{j=n+1}^N \mathbb{1}_{\{u_j\in B_{\mathbb{H}}(u_i)\}}$. Thus

$$\begin{aligned}&\mathbb{E}_n\left[\sum_{i=1}^n \mathbb{1}_{\{\deg_n(u_i)=k_n\}}\mathbb{1}_{\{|B_{\mathbb{H}}(u_i)\cap W|\geq 1\}}\mathbb{1}_{Q_n(u_i)}\right]\\ &\leq \sum_{i=1}^n\sum_{j=n+1}^N \mathbb{E}_n\left[\mathbb{1}_{\{\deg_n(u_i)=k_n\}}\mathbb{1}_{Q_n(u_i)}\mathbb{1}_{\{u_j\in B_{\mathbb{H}}(u_i)\}}\right]\\ &= n(N-n)\mathbb{P}_n\left(\mathrm{Bin}\left(n-2,\frac{\mu_{\mathbb{H}}(U_1)}{n}\right)=k_n, Q_n(U_1), U_2\in B_{\mathbb{H}}(U_1)\right)\\ &= n(N-n)\mathbb{E}_n\left[\mathbb{1}_{Q_n(U_1)}\mathbb{P}\left(\mathrm{Bin}\left(n-2,\frac{\mu_{\mathbb{H}}(U_1)}{n}\right)=k_n, U_2\in B_{\mathbb{H}}(U_1)\middle| U_1\right)\right]\\ &= n(N-n)\mathbb{E}_n\left[\mathbb{1}_{Q_n(U_1)}\mathbb{P}\left(\mathrm{Bin}\left(n-2,\frac{\mu_{\mathbb{H}}(U_1)}{n}\right)=k_n\middle| U_1\right)\frac{\mu_{\mathbb{H}}(U_1)}{n}\right],\end{aligned}$$

where the last step follows since, conditionally on U_1, the events $\mathrm{Bin}\left(n-2,\frac{\mu_{\mathbb{H}}(U_1)}{n}\right) = k_n$ and $U_2 \in B_{\mathbb{H}}(U_1)$ are independent and $\mathbb{P}\left(U_2\in B_{\mathbb{H}}(U_1)\middle| U_1\right) = \mu_{\mathbb{H}}(U_1)/n$. In addition, on the event $Q_n(U_1)$ we have that $\mu_{\mathbb{H}}(U_1) \le k_n + C\sqrt{k_n \log(k_n)}$ and thus the second statement of Lemma D.1 implies that

$$\begin{aligned}
&\mathbb{E}_n\left[\mathbb{1}_{\{Q_n(U_1)\}}\mathbb{P}\left(\operatorname{Bin}\left(n-2, \frac{\mu_{\mathbb{H}}(U_1)}{n}\right) = k_n \middle| U_1\right)\right] \\
&= (1+o(1))\mathbb{E}_n\left[\mathbb{1}_{\{Q_n(U_1)\}}\mathbb{P}\left(\operatorname{Bin}\left(N-1, \frac{\mu_{\mathbb{H}}(U_1)}{n}\right) = k_n \middle| U_1\right)\right] \\
&\le O(1)\,\mathbb{P}_n\left(\operatorname{Bin}\left(N-1, \frac{\mu_{\mathbb{H}}(U_1)}{n}\right) = k_n\right) \\
&= O(1)\,\mathbb{P}_n\left(\deg_N(U_1) = k_n\right).
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
&\mathbb{E}_n\left[\sum_{i=1}^n \mathbb{1}_{\{\deg_n(u_i)=k_n\}}\mathbb{1}_{\{|B_{\mathbb{H}}(u_i)\cap W|\ge 1\}}\mathbb{1}_{Q_n(u_i)}\right] \\
&\qquad\le (N-n)\left(k_n + C\sqrt{k_n\log(k_n)}\right)\mathbb{P}_n\left(\deg_N(U_1) = k_n\right) \\
&\qquad= (1+o(1))\sqrt{n\log(n)}k_n\mathbb{P}_n\left(\deg_N(U_1) = k_n\right) \\
&\qquad= o\left(n\mathbb{P}_n\left(\deg_N(U_1) = k_n\right)\right) = o\left(\mathbb{E}_n\left[N_{\mathrm{Po}}(k_n)\right]\right),
\end{aligned}$$

where we have used again that $\sqrt{n\log(n)}k_n = o(n)$. □

With this result we can now prove Proposition 4.3.3, which states

$$\lim_{n\to\infty} s(k_n)\mathbb{E}\left[|c^*(k_n; G_n) - c^*(k_n; G_{\mathrm{Po}})|\right] = 0.$$

Proof of Proposition 4.3.3. First we note that Proposition 4.3.4, 4.3.5 and 4.3.6 together imply that

$$\mathbb{E}\left[c^*(k_n; G_{\mathrm{Po}})\right] = (1+o(1))s(k_n)$$

Therefore it suffices to show that

$$\mathbb{E}\left[|c^*(k_n; G_n) - c^*(k_n; G_{\mathrm{Po}})|\right] = o\left(\mathbb{E}\left[c^*(k_n; G_{\mathrm{Po}})\right]\right).$$

For this we observe that we are looking at the modified clustering coefficient, where we divide by the expected number of degree k_n vertices. As the expected numbers of degree k_n vertices in G_{Po} and G_n are asymptotically equivalent (see Lemma 4.7.5), it is therefore sufficient to consider the sum of the clustering coefficients of all vertices of degree k_n. Given again the standard coupling between the binomial and Poisson process (as used in the proof of Lemma 4.7.5), we again denote by $\mathcal{V}_n(k_n)$ the set of degree k_n vertices in G_n and by $\mathcal{V}_{\mathrm{Po}}(k_n)$ the set of degree k_n vertices in G_{Po}. If a vertex is contained in both sets, then it must have the same degree in both the Poisson and KPKVB graph, and given the nature of the coupling, the neighbourhoods are therefore the same and hence also their clustering coefficients agree.

The difference of the sum of the clustering coefficients therefore comes from all the clustering coefficients of the symmetric difference $\mathcal{V}_n(k_n)\Delta\mathcal{V}_{\mathrm{Po}}(k_n)$. By Lemma 4.7.5 the expected number vertices in this set is $\mathbb{E}\left[|N_n(k_n) - N_{\mathrm{Po}}(k_n)|\right] = o\left(\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]\right)$. Therefore,

$$\begin{aligned}\mathbb{E}\left[|c^*(k_n; G_n) - c^*(k_n; G_{\mathrm{Po}})|\right] &\leq \frac{\mathbb{E}\left[|N_n(k_n) - N_{\mathrm{Po}}(k_n)|\right]}{(1 + o\left(1\right))\mathbb{E}\left[N_{\mathrm{Po}}(k_n)\right]}\mathbb{E}\left[c^*(k_n; G_{\mathrm{Po}})\right] \\ &= o\left(1\right)\mathbb{E}\left[c^*(k_n; G_{\mathrm{Po}})\right],\end{aligned}$$

which finishes the proof. □

Finally we prove Lemma 4.3.2, whose statement is

$$|c^*(k_n; G_n) - c(k_n; G_n)| = o_{\mathbb{P}}\left(s(k_n)\right).$$

Proof of Lemma 4.3.2. Since Propositions 4.3.3-4.3.6 imply that

$$\mathbb{E}\left[c^*(k_n; G_n)\right] = O\left(s(k_n)\right),$$

and since

$$\frac{|N_n(k_n) - \mathbb{E}\left[N_n(k_n)\right]|}{N_n(k_n)} = o_{\mathbb{P}}\left(1\right),$$

we immediately infer that

$$|c^*(k_n; G_n) - c(k_n; G_n)| = c^*(k_n; G_n)\left|\frac{\mathbb{E}\left[N_n(k_n)\right]}{N_n(k_n)} - 1\right| = o_{\mathbb{P}}\left(1\right).$$

□

Summary

In this thesis, we study a random graph model proposed by Krioukov et al. [36] in 2010. In this model, vertices are chosen randomly inside a disk around the origin in the hyperbolic plane and two vertices are adjacent if their hyperbolic distance is at most the radius of the disk. The model is specified using three parameters: the number of vertices n, which we think of as going to infinity, and $\alpha, \nu > 0$, which we think of as constant. The parameter α determines how we distribute the points in the disk. Roughly speaking, the larger α, the more points will be close to the boundary of the disk. The parameter ν is used in the definition of the radius of the disk in such a way that the average degree tends to a finite positive constant with high probability as $n \to \infty$ for $\alpha > \frac{1}{2}$.

Since its invention by Krioukov et al. in 2010, this model has attracted significant attention as a promising model for real-world networks (like the Internet or social networks). It has been shown that, for $\alpha > \frac{1}{2}$, this model simultaneously satisfies many properties that were identified as common in a wide range of real-world networks by previous research in networks science. These properties include sparsity (constant average degree), a power-law degree distribution (up to a certain scaling), 'short distances', a non-vanishing clustering coefficient and the existence of a giant component.

The first graph-theoretical concepts considered in this thesis are perfect matchings and Hamilton cycles. As the model has been shown to have (a linear fraction of) isolated vertices for $\alpha > \frac{1}{2}$, there cannot be any perfect matchings or Hamilton cycles for $\alpha > \frac{1}{2}$. We show that for $\alpha < \frac{1}{2}$, there is a non-trivial phase transition in ν for both of these properties, i.e. for every $\alpha < 1/2$ and $\nu = \nu(\alpha)$ sufficiently small, the model does not contain a perfect matching or Hamilton cycle with high probability, whereas for every $\alpha < 1/2$ and $\nu = \nu(\alpha)$ sufficiently large, the model contains a Hamilton cycle (and hence also a perfect matching) with high probability.

Secondly, it has previously been shown that the model has a power-law with exponent $2\alpha + 1$ for $\alpha > \frac{1}{2}$, more precisely, that the fraction of vertices of degree k_n is proportional to $k_n^{-(2\alpha+1)}$ with high probability as $n \to \infty$ for sequences k_n which are bounded above by a certain asymptotic scaling. We improve upon this by showing that the degree distribution follows a power-law with exponent $2\alpha + 1$

up to the maximum scaling of $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$, establishing a Poisson limit law in the boundary case that $k_n = (1+o(1))cn^{\frac{1}{2\alpha+1}}$ for some fixed c and showing that there are indeed no vertices of degree exactly k_n with high probability if k_n grows asymptotically faster than $n^{\frac{1}{2\alpha+1}}$.

Thirdly, we show that for $\alpha > \frac{1}{2}$, the clustering coefficient tends in probability to a constant γ that we give explicitly as a closed-form expression in terms of α, ν and certain special functions. This improves earlier work by Gugelmann et al. [30], who proved that the clustering coefficient remains bounded away from zero with high probability, but left open the issue of convergence to a limiting constant. Similarly, we are able to show that $c(k)$, the average clustering coefficient over all vertices of degree exactly k, tends in probability to a limit $\gamma(k)$ which we can give explicitly as a closed-form expression in terms of α, ν and certain special functions. We are able to extend this last result also to sequences $(k_n)_n$ where k_n grows as a function of n. Our results show that $\gamma(k)$ scales differently, as k grows, for different ranges of α. More precisely, $\gamma(k) = \Theta(k^{2-4\alpha})$ if $\frac{1}{2} < \alpha < \frac{3}{4}$, $\gamma(k) = \Theta(\log(k)/k)$ if $\alpha = \frac{3}{4}$ and $\gamma = \Theta\left(k^{-1}\right)$ if $\alpha > \frac{3}{4}$. In each case, we also determine the leading constant. These results (partially) contradict a claim of Krioukov et al., which stated that the limiting values $\gamma(k)$ should always scale with k^{-1} as we let k grow (irregardless of the value of α). We perform simulations which confirm the limiting values γ and $\gamma(k)$.

Samenvatting

In dit proefschrift bestuderen we een model van stochastische grafen en complexe netwerken voorgesteld door Krioukov et al. [36] in 2010. In dit model worden knopen willekeurig gekozen binnen een schijf in het hyperbolische vlak en twee knopen zijn aangrenzend als ze maximaal een bepaalde hyperbolische afstand van elkaar verwijderd zijn. Het model wordt gespecificeerd met behulp van drie parameters: het aantal knopen n, waarvan we aannemen dat die naar oneindig gaan, en twee positieve reële getallen $\alpha, \nu > 0$, die we constant nemen. Grof gezegd, hoe groter α, hoe meer punten dichtbij de grens van de schijf zullen zitten. De parameter ν wordt gebruikt in de definitie van de straal van de schijf zodanig dat de gemiddelde graad met een hoge waarschijnlijkheid neigt naar een positieve constante als $n \to \infty$ voor $\alpha > \frac{1}{2}$.

Sinds zijn uitvinding door Krioukov et al. in 2010 heeft dit model veel aandacht gekregen als een veelbelovend model voor netwerken (zoals het internet of sociale netwerken). Het is aangetoond dat dit model voor $\alpha > \frac{1}{2}$ aan veel eigenschappen voldoet die werden geïdentificeerd als gebruikelijk in een breed scala van netwerken door eerder onderzoek in netwerkwetenschap. Deze eigenschappen zijn onder meer een constante gemiddelde graad, een machtsverdeling van de graden (tot een bepaalde schaal), 'korte afstanden', een niet-verdwijnende clusteringcoëfficiënt en het bestaan van een groot verbonden component.

De eerste grafentheoretische concepten die beschouwd worden in dit proefschrift, zijn perfecte koppelingen en Hamiltoncykel. Omdat het is aangetoond dat het model (een lineaire fractie van) geïsoleerde knopen heeft, kunnen er geen perfecte koppelingen of Hamiltoncykel zijn voor $\alpha > \frac{1}{2}$. Wij laten zien dat er voor $\alpha < \frac{1}{2}$ een niet-triviale faseovergang voor beide eigenschappen is. Dat wil zeggen dat voor elke $\alpha < \frac{1}{2}$ en $\nu = \nu(\alpha)$ voldoende klein, het model met hoge waarschijnlijkheid geen perfecte koppeling of Hamiltoncykel bevat, terwijl voor elke $\alpha < \frac{1}{2}$ en $\nu = \nu(\alpha)$ voldoende groot, het model met grote waarschijnlijkheid een Hamiltoncykel (en dus ook een perfecte koppeling) bevat.

Ten tweede is eerder aangetoond dat de verdeling van de graden in dit model een machtswet heeft met exponent $2\alpha + 1$ voor $\alpha > \frac{1}{2}$. Dit betekend dat de fractie van knopen van graad k_n evenredig is met $k_n^{-(2\alpha+1)}$ met grote waarschijnlijkheid als $n \to \infty$ voor alle rijen k_n die naar boven worden begrensd door een bepaalde asymptotische schaling. We verbeteren dit resultaat door aan te tonen dat deze machtswet met exponent $2\alpha + 1$ tot de maximale schaal van $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$ geldt. Ook laten we zien dat er een Poisson limietwet bestaat in het grensgeval dat $k_n = (1 + o(1))cn^{\frac{1}{2\alpha+1}}$ voor een vaste c en dat er inderdaad geen knopen zijn die precies graad k_n hebben met grote waarschijnlijkheid als k_n asymptotisch sneller dan $n^{\frac{1}{2\alpha+1}}$ groeit.

Ten derde laten we zien dat de clusteringcoëfficiënt voor $\alpha > \frac{1}{2}$ in waarschijnlijkheid neigt naar een constante γ die we expliciet geven als een gesloten vorm in termen van α, ν en bepaalde speciale functies. Dit is een verbetering ten opzichte van eerder werk van Gugelmann et al. [30], die bewezen dat de clusteringcoëfficiënt met grote waarschijnlijkheid groter dan nul blijft, maar de kwestie van convergentie open lieten. Ook kunnen we aantonen dat $c(k)$, de gemiddelde clusteringcoëfficiënt over alle knopen van graad exact k, convergeert naar een functie $\gamma(k)$. Voor deze limiet geven we ook een expliciete expressie in gesloten vorm in termen van α, ν en bepaalde speciale functies. We kunnen dit laatste resultaat ook uitbreiden naar rijen $(k_n)_n$ waar k_n groeit als functie van n. Onze resultaten laten zien dat $\gamma(k)$ anders schaalt, naarmate k groeit, voor verschillende bereiken van α. Om precies te zijn, $\gamma(k) = \Theta(k^{2-4\alpha})$ als $\frac{1}{2} < \alpha < \frac{3}{4}$, $\gamma(k) = \Theta(\log(k)/k)$ als $\alpha = \frac{3}{4}$ en $\gamma(k) = \Theta\left(k^{-1}\right)$ als $\alpha > \frac{3}{4}$. Deze resultaten zijn in tegenspraak met een bewering van Krioukov et al., waarin werd gesteld dat de grenswaarden $\gamma(k)$ altijd moeten worden geschaald met k^{-1} als we k laten groeien. We voeren simulaties uit die de grenswaarden γ en $\gamma(k)$ bevestigen.

Zusammenfassung

In der vorliegenden Doktorarbeit untersuchen wir ein von Krioukov et al. [36] im Jahr 2010 vorgeschlagenes Zufallsgraphenmodell. In diesem Modell werden innerhalb einer Scheibe um den Ursprung der hyperbolischen Ebene Knoten zufällig verteilt; zwei Knoten heißen benachbart, wenn ihr hyperbolischer Abstand höchstens dem Radius der Scheibe entspricht. Das Modell wird mit drei Parametern spezifiziert: der Anzahl der Knoten n, die wir gegen Unendlich streben lassen, und zwei reelle Zahlen $\alpha, \nu > 0$, die wir als konstant betrachten. Der Parameter α bestimmt, wie die Knoten innerhalb der Scheibe verteilt werden. Vereinfacht gesagt: je größer α, desto mehr Punkte befinden sich nahe der Grenze der Scheibe. Der Parameter ν fließt bei der Definition des Radius der Scheibe derart ein, dass der durchschnittliche Grad für $\alpha > \frac{1}{2}$ in Wahrscheinlichkeit gegen eine endliche positive Konstante konvergiert, wenn $n \to \infty$.

Seit seiner Erfindung durch Krioukov et al. 2010 hat dieses Modell erhebliche Aufmerksamkeit auf sich gezogen und wird als vielversprechendes Modell für reale Netzwerke (wie das Internet oder soziale Netzwerke) gehandelt. Es wurde gezeigt, dass dieses Modell für $\alpha > \frac{1}{2}$ gleichzeitig mehrere Eigenschaften erfüllt, die in früheren Untersuchungen der Netzwerkwissenschaft als Kennzeichen vieler realer Netzwerke identifiziert wurden. Diese Eigenschaften umfassen Sparsity (konstanter Durchschnittsgrad), eine Potenzgesetz-Gradverteilung (bis zu einer bestimmten Skalierung der Grade), "kurze Entfernungen", einen nicht verschwindenden Clustering-Koeffizienten und die Existenz einer riesigen Komponente (engl. giant component).

Die ersten graphentheoretischen Konzepte, die in dieser Arbeit betrachtet werden, sind perfekte Matchings (gelegentlich auch perfekte Paarungen genannt) und Hamilton-Zyklen. Da gezeigt wurde, dass das Modell für $\alpha > \frac{1}{2}$ (einen linearen Anteil) isolierter Knoten aufweist, kann es für $\alpha > \frac{1}{2}$ keine perfekten Matchings oder Hamilton-Zyklen geben. Wir zeigen, dass es für $\alpha < \frac{1}{2}$ für beide Eigenschaften einen nicht-trivialen Phasenübergang in ν gibt, d.h. für jedes $\alpha < \frac{1}{2}$ und $\nu = \nu(\alpha)$ ausreichend klein, enthält das Modell mit hoher Wahrscheinlichkeit keine perfekten Matchings- oder Hamilton-Zyklen, während für jedes $\alpha < \frac{1}{2}$ und $\nu = \nu(\alpha)$ ausreichend groß, mit hoher Wahrscheinlichkeit ein Hamilton-Zyklus (und damit auch ein perfektes Matching) im Modell enthalten ist.

Zweitens wurde es bereits gezeigt, dass das Modell für $\alpha > \frac{1}{2}$ einem Potenzgesetz mit Exponent $2\alpha + 1$ folgt, genauer gesagt, dass mit hoher Wahrscheinlichkeit der Anteil der Knoten des Grades k_n proportional zu $k_n^{-(2\alpha+1)}$ ist, wenn $n \to \infty$ für alle Folgen k_n, die nach oben durch eine bestimmte asymptotische Skalierung begrenzt sind. Wir verbessern dies, indem wir zeigen, dass das Potenzgesetz mit Exponent $2\alpha + 1$ für die Gradverteilung bis zur maximalen Skalierung von $k_n = o\left(n^{\frac{1}{2\alpha+1}}\right)$ gilt. Wir beweisen ein Poisson-Grenzgesetz für den Fall, dass $k_n = (1+o(1))cn^{\frac{1}{2\alpha+1}}$ für ein festes c und wir zeigen, dass es mit hoher Wahrscheinlichkeit tatsächlich keine Knoten mit Grad genau k_n gibt, wenn k_n asymptotisch schneller wächst als $n^{\frac{1}{2\alpha+1}}$.

Drittens zeigen wir, dass für $\alpha > \frac{1}{2}$ der Clustering-Koeffizient in Wahrscheinlichkeit gegen eine Konstante γ konvergiert, die wir explizit als geschlossenen Ausdruck in α, ν und speziellen Funktionen angeben. Dies verbessert frühere Arbeiten von Gugelmann et al. [30], die bewiesen, dass der Clustering-Koeffizient mit hoher Wahrscheinlichkeit echt größer Null ist, aber das Problem der Konvergenz offen ließen. In ähnlicher Weise können wir zeigen, dass $c(k)$, der durchschnittliche Clustering-Koeffizient über alle Knoten mit Grad genau k, in Wahrscheinlichkeit gegen eine Konstante $\gamma(k)$ konvergiert, die wir explizit als geschlossenen Ausdruck in α, ν und speziellen Funktionen angeben. Wir können dieses letzte Ergebnis auch auf Folgen $(k_n)_n$ ausdehnen, für die k_n als Funktion von n wächst. Unsere Ergebnisse zeigen, dass $\gamma(k)$ für verschiedene Werte von α unterschiedlich skaliert, wenn k gegen Unendlich tendiert. Genauer gesagt ist $\gamma(k) = \Theta(k^{2-4\alpha})$, für $\frac{1}{2} < \alpha < \frac{3}{4}$, $\gamma(k) = \Theta(\log(k)/k)$ für $\alpha = \frac{3}{4}$ und $\gamma = \Theta\left(k^{-1}\right)$ für $\alpha > \frac{3}{4}$. Für jeden Fall bestimmen wir auch die Leitkonstante. Diese Ergebnisse widersprechen (teilweise) einer Behauptung von Krioukov et al., die besagte, dass die Grenzwerte $\gamma(k)$ immer mit k^{-1} skalieren, wenn $k \to \infty$ (unabhängig vom Wert von α). Wir führen Simulationen durch, die die Grenzwerte γ und $\gamma(k)$ bestätigen.

Acknowledgements

I am very grateful for all the people and circumstances that allowed me, directly or indirectly, to write and complete this doctoral thesis.

I would like to thank Universiteit Utrecht, Rijksuniversiteit Groningen and NWO for funding my PhD position and research. I would like to thank all sources of financial support I received for attending mathematical schools, workshops and conferences. I would like to thank all sources of financial support for trips devoted to research collaborations.

I would like to thank my supervisor Prof Tobias Müller, most of all for supporting my development in becoming an independent mathematician. In particular, I benefitted from his questions when presenting mathematical background papers or theory, the mathematical discussions in his office and his feedback on my mathematical writing with emphasis on quality and rigor. Furthermore, he encouraged and supported my participation in an extraordinary number of international mathematical schools, workshops, conferences and research-related trips. I would like to thank my supervisors Prof Roberto Fernandez in Utrecht and Prof Ernst Wit in Groningen.

I would like to thank my co-authors Dr Nikolaos Fountoulakis, in particular for our discussions during my visit in Birmingham, Dr Pim van der Hoorn, in particular for the many meetings both in person and via Skype and his stimulating emails, and Dr Dieter Mitsche, in particular for the discussions during his visit in Groningen. I would like to thank all my co-authors for adding a social dimension to my mathematical work and all the benefits that come with it.

I would like to thank the members of the assessment committee Prof Remco van der Hofstad, Prof Konstantinos Panagiotou and Prof Daniel Valesin for reading and evaluating the manuscript of this thesis and providing valuable feedback.

I am very grateful for the opportunity to take part in an extraordinary number of international mathematical schools, workshops and conferences where I learned a lot of new mathematics which will hopefully become useful in my future career - e.g. discrete Fourier analysis in Barcelona (Spain), random graphs and processes in Toronto (Canada), Gaussian free fields in Haifa (Israel), stochastic partial differential equations in Klosterneuburg (Austria) or the travelling salesman problem and sample complexity in Valparaiso (Chile) - or was able to present my work, either as a poster or in a talk - e.g. at RSA in Zurich (Switzerland).

I would like to thank my teachers and professors who prepared me before my PhD. I would like to thank the supervisor of my master's thesis Prof Ben Green, my master's career advisor Prof Minhyong Kim and my master's college advisor Prof Mason Porter. I would like to thank the supervisor of my bachelor's thesis Prof Andreas Gathmann. I would like to thank my teachers, in particular Angelika Diedrich, Nicolas Frank, Monika Goodson, Elisabeth Haas, Christoph Holtwiesche, Margit Kaschta, Dr. Heinz-Martin Müller and Werner Ostendorf.

Among my PhD friends, I would like to thank Benjamin Hansen, in particular for accompanying my PhD journey from the beginning like a brother, Réka Szabó, e.g. for organizing social events and sharing her experiences with the PhD procedure, Gabriel Baptista, e.g. for playing music together and talking about interesting books, and Teun Verstraaten, e.g. for helping me improve my Dutch, in Groningen, I would like to thank Lorenzo Amabili, Deepan Basu, Matthijs Ebbens, Krisztina Kruzslicz, Bohuan Lin, Nikolay Martynchuk, Georg Wilding, Yongjiao Zhang, in Utrecht, I would like to thank Felix Beckebanze.

I would like to thank all my long-time friends, in particular Teemu Belt, Niklas Bockius, Eric Breuer, Alicia Chew, Nicolas Dietrich, Tom Ewen, Martin Gockert, Jin Wen Goh, Michael Hedderich, Julia Ichihara, Dominik Kirst, Yibei Li, Zbynek Loebl, Shivani Malik, Steve Nugent, Moumanti Podder, Bernhard Scheck, Markus Schmidtchen, Myriam Shafie, Elisabeth Steinbrecher, Luc Stultiens, Vaanathi Sundaresan and Richard Vogg (and so many more that I cannot mention here).

I would like to thank my family, in particular my parents Cornelia and Christoph Schepers, my grandparents Erika and Hans Reis, Elisabeth and Theodor Schepers and my brother Adrian Schepers. My deep gratitude for my family's love cannot be expressed in words.


Merci!
谢谢!
¡Gracias!
Спасибо!
Köszönöm!
Dank u wel!
Danke!
Kiitos!
Gratias tibi!
Ευχαριστώ!
ありがとう！
Thank you!
धन्यवाद!
תודה!
Dziękuję!
Obrigado!
감사합니다!
Grazie!
Děkuji!
Danke schön!
Dongkschää!

Bibliography

[1] *Nist digital library of mathematical functions*, Release 1.0.22 of 2019-03-15, F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[2] M.A. Abdullah, M. Bode, and N. Fountoulakis, *Typical distances in a geometric model for complex networks*, Internet Mathematics **1** (2017).

[3] M. Ajtai, J. Komlós, and E. Szemerédi, *First occurrence of Hamilton cycles in random graphs*, North-Holland Mathematics Studies **115** (1985), 173–175.

[4] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74** (2002), no. 1, 47–97.

[5] D. Aldous and J. M. Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on Discrete Structures (2004), 1–72.

[6] Noga Alon and Joel Spencer, *The probabilistic method*, Wiley, 2015.

[7] George E Andrews, Richard Askey, and Ranjan Roy, *Special functions*, vol. 71, Cambridge university press, 2000.

[8] J. Balogh, B. Bollobás, M. Krivelevich, T. Müller, and M. Walters, *Hamilton cycles in random geometric graphs*, The Annals of Applied Probability **21** (2011), no. 3, 1053–1072.

[9] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab., 6, no. 23, 13 pp. (electronic) (2001).

[10] Jean-Claude Bermond, *Hamiltonian graphs. selected topics in graph theory*, (1979), chapter 6, pp.127–167.

[11] M. Bode, N. Fountoulakis, and T. Müller, *On the largest component of a hyperbolic model of complex networks*, Electronic Journal of Combinatorics **22** (2015), no. 3, Paper P3.24, 43 pages.

[12] ———, *The probability that the hyperbolic random graph is connected*, Random Structures Algorithms **49** (2016), no. 1, 65–94.

[13] B. Bollobás, *The evolution of sparse graphs*, Graph Theory and Combinatorics (London), Academic Press, 1984, pp. 35–57.

[14] Janos Bolyai, *Appendix, scientiam spatii absolute veram exhibens: a veritate aut falsitate axiomatis xi. euclidei (a priori haud unquam decidenda) independentem; adjecta ad casum falsitatis, quadratura circuli geometrica. auctore johanne bolyai de eadem, geometrarum in exercitu caesareo regio austriaco castrensium capitaneo.*, Coll. Ref., 1832.

[15] R. Bonola, *Non-euclidean geometry: A critical and historical study of its development*, Chicago: Open Court, 1912.

[16] Daniel Callahan and John Casey, *Euclid's "elements" redux*, 2015.

[17] Elisabetta Candellero and Nikolaos Fountoulakis, *Clustering and the hyperbolic geometry of complex networks*, Internet Mathematics **12** (2016), no. 1-2, 2–53.

[18] Hanshuang Chen, Chuansheng Shen, Gang He, Haifeng Zhang, and Zhonghuai Hou, *Critical noise of majority-vote model on complex networks*, Physical Review E **91** (2015), no. 2, 022816.

[19] Kimberly Claffy, Young Hyun, Ken Keys, Marina Fomenkov, and Dmitri Krioukov, *Internet mapping: from art to science*, Conference For Homeland Security, 2009. CATCH'09. Cybersecurity Applications & Technology, IEEE, 2009, pp. 205–211.

[20] Brian Davies, *Integral transforms and their applications*, vol. 41, Springer Science & Business Media, 2012.

[21] J. Díaz, D. Mitsche, and X. Pérez, *Sharp threshold for Hamiltonicity of random geometric graphs*, SIAM J. Discrete Math. **21** (2007), no. 1, 57–65.

[22] Manfredo P. do Carmo, *Differential geometry of curves and surfaces.*, Prentice Hall, 1976.

[23] Arthur Erdélyi, Wilhelm Magnus, Fritz Oberhettinger, Francesco G Tricomi, et al., *Higher transcendental functions, vol. 1*, 1953.

[24] N. Fountoulakis, *On the evolution of random graphs on spaces with negative curvature*, ArXiv e-prints (2012).

[25] Nikolaos Fountoulakis and Tobias Müller, *Law of large numbers for the largest component in a hyperbolic model of complex networks*, The Annals of Applied Probability **28** (2018), no. 1, 607–650.

[26] T. Friedrich and A. Krohmer, *On the diameter of hyperbolic random graphs*, SIAM J. Disc. Math. **32** (2018), 1314–1334.

[27] A. Frieze, X. Pérez-Giménez, P. Prałat, and B. Reiniger, *Perfect matchings and hamiltonian cycles in the preferential attachment model*, Random Structures and Algorithms, to appear.

[28] Izrail Solomonovich Gradshteyn and Iosif Moiseevich Ryzhik, *Table of integrals, series, and products*, Elsevier, 2015.

[29] L. Gugelmann, K. Panagiotou, and U. Peter, *Random hyperbolic graphs: degree sequence and clustering*, Preprint. Available from http://arxiv.org/abs/1205.1470. Conference version in ICALP 2012.

[30] ______, *Random hyperbolic graphs: Degree sequence and clustering*, Proceedings of the 39th International Colloquium Conference on Automata, Languages, and Programming - Volume Part II (Berlin, Heidelberg), ICALP'12, Springer-Verlag, 2012, pp. 573–585.

[31] M. Kiwi and D. Mitsche, *Spectral gap of random hyperbolic graphs and related parameters*, Annals of Applied Probability **28** (2018), 941–989.

[32] M. A. Kiwi and D. Mitsche, *A bound for the diameter of random hyperbolic graphs*, Proceedings of the Twelfth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2015, San Diego, CA, USA, January 4, 2015 (Robert Sedgewick and Mark Daniel Ward, eds.), SIAM, 2015, pp. 26–39.

[33] ______, *On the second largest component of random hyperbolic graphs*, Available at `arXiv:1712.02828v1`, 2017.

[34] J. Komlós and E. Szemerédi, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), 55–63.

[35] A.D. Korshunov, *A solution of a problem of P. Erdős and A. Rényi about Hamilton cycles in non-oriented graphs*, Metody Diskr. Anal. Teoriy Upr. Syst., Sb. Trudov Novosibirsk **31** (1977), 17–56 (in Russian).

[36] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, *Hyperbolic geometry of complex networks*, Phys. Rev. E (3) **82** (2010), no. 3, 036106, 18. MR 2787998 (2012a:05286)

[37] G. Last and M. Penrose, *Lectures on the Poisson process*, IMS Textbooks, Cambridge University Press, 2018.

[38] Yudell L Luke, *Mathematical functions and their approximations*, Academic Press, 2014.

[39] Cornelis Simon Meijer, *On the G-function. I-VIII*, Nederl. Akad. Wetensch., Proc., vol. 49, 1946.

[40] M Carmen Miguel, Jack T Parley, and Romualdo Pastor-Satorras, *Effects of heterogeneous social interactions on flocking dynamics*, Physical review letters **120** (2018), no. 6, 068303.

[41] T. Müller, X. Pérez-Giménez, and N.C. Wormald, *Disjoint hamilton cycles in the random geometric graph*, Journal of Graph Theory **68** (2011), no. 4, 299–322.

[42] T. Müller and M. Staps, *The diameter of KPKVB random graphs*, Available at `arXiv 1707.09555`, 2017.

[43] translated by George Bruce Halsted Nicholas Lobachevski, *Geometrical researches on the theory of parallels*, Open Court Publishing Company, 1891.

[44] M. D. Penrose, *Random geometric graphs*, Oxford Studies in Probability, vol. 5, Oxford University Press, Oxford, 2003. MR 1986198 (2005j:60003)

[45] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364.

[46] R.W. Robinson and N.C. Wormald, *Almost all regular graphs are hamiltonian*, Random Structures and Algorithms **5** (1994), 363–374.

[47] Clara Stegehuis, Remco van der Hofstad, and Johan SH van Leeuwaarden, *Scale-free network clustering in hyperbolic and other random graphs*, Journal of Physics A: Mathematical and Theoretical (2019).

[48] J. Stillwell, *Geometry of surfaces*, Universitext, Springer-Verlag, New York, Berlin, Heidelberg, 1992.

[49] Nico M Temme, *Special functions: An introduction to the classical functions of mathematical physics*, John Wiley & Sons, 2011.

[50] R. van der Hofstad, *Random graphs and complex networks*, vol. 2, 2018+, in preparation. Available at http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf.

List of notation

$G_n = G(n;\alpha,\nu)$	KPKVB model
$G_{Po} = G_{Po}(n;\alpha,\nu)$	poissonized KPKVB model
$\overline{G_{Po}}$	poissonized KPKVB model in coordinates of box model
$G_{box} = G_{box}(n;\alpha,\nu)$	box model
$G_{box,H}$	box model truncated at height H
$G_\infty = G_\infty(\alpha,\nu)$	infinite (limit) model
$R = R_n$	$2\ln\frac{n}{\nu}$ radius of hyperbolic disk and adjacency threshold
$\mathbb{H}$	hyperbolic plane with curvature -1
$\mathcal{D}$ or $\mathcal{D}_R$	disk of radius R around origin of hyperbolic plane $\mathbb{H}$
I_n	$\frac{\pi}{2}e^{\frac{R}{2}}$
$\mathcal{R}$	box of the box model, i.e. $= (-I_n, I_n] \times (0, R] \subset \mathbb{R}^2$
$\mathcal{R}(I)$	$(-I_n, I_n] \times I$ for $I \subset [0, R]$
$\mathcal{R}_H = \mathcal{R}([0,H])$	box truncated at height H
Ψ	coordinate transformation given by $\Psi : [0,R] \times (-\pi,\pi] \to \mathcal{R}$, $\Psi(r,\theta) = \left(\theta\frac{e^{R/2}}{2}, R - r\right)$
$\vartheta(r,r')$	maximal angle between two adjacent vertices with radial coordinates r, r' in KPKVB model, i.e. $\vartheta(r,r') = \begin{cases} \arccos\left(\frac{\cosh r \cosh r' - \cosh R}{\sinh r \sinh r'}\right), & \text{if } r + r' \geq R, \\ \pi, & \text{if } r + r' < R. \end{cases}$
$\Phi(y,y')$	$\Phi(y,y') = \frac{1}{2}e^{R/2}\vartheta(R-y, R-y')$.
$\mathcal{B}_\infty(p)$	neighbourhood ball of $p \in \mathbb{R} \times (0,\infty)$ in infinite model, i.e. $\{p' \in \mathbb{R} \times (0,\infty) : \lvert x - x'\rvert \leq e^{\frac{y+y'}{2}}\}$
$\mathcal{B}_{\text{box}}(p)$	neighbourhood ball of $p \in \mathcal{R}$ in finite box model, i.e. $\{p' \in \mathcal{R} : \lvert x - x'\rvert \leq e^{\frac{y+y'}{2}}\}$
$\mathcal{B}(p)$	neighbourhood ball of $p = (x,y) \in \mathcal{R}$ which is induced by hyperbolic metric, i.e. $\Psi(\{u \in \mathcal{D} : d_{\mathbb{H}}(\Psi^{-1}(p), u) \leq R\})$ $= \{p' = (x',y') \in \mathcal{R} : \lvert x - x'\rvert_{\pi e^{\frac{R}{2}}} \leq \Phi(y,y')\} \subset \mathcal{R}$

$f(x,y) = f_{\alpha,\nu}(x,y)$	$\frac{\alpha\nu}{\pi} e^{-\alpha y} \cdot \mathbb{1}_{\{y>0\}}$
$f_n(x,y)$ or $f_{n,\alpha,\nu}(x,y)$	$\frac{\alpha\nu}{\pi} e^{-\alpha y} \cdot \mathbb{1}_{\{-\frac{\pi n}{2\nu} < x \leq \frac{\pi n}{2\nu}, 0<y<2\ln\frac{n}{\nu}\}}$
$g(r,\theta) = g_{\alpha,R}(r,\theta)$	$\frac{\alpha\sinh(\alpha r)}{2\pi(\cosh(\alpha R)-1)} \mathbb{1}_{\{0\leq r\leq R, -\pi<\theta\leq\pi\}}$
$\mu = \mu_{\alpha,\nu}$	measure with density function f i.e. for every Borel-measurable subset $S \subset \mathbb{R}^2$ we have $\mu(S) = \int_S f(x,y)\,dx\,dy$
μ_n	measure with density function f_n
μ_g	measure with density function g
$\mathcal{P} = \mathcal{P}_{\alpha,\nu}$	vertex set of infinite limit model, i.e. Poisson point process on $\mathbb{R}^2$ with intensity function f, resp. intensity measure μ
$\mathcal{V}_n$ or $\mathcal{V}_{\text{box}}$	vertex set of box model, i.e. Poisson point process with intensity function f_n, resp. intensity measure μ_n, equivalently it is given by $\mathcal{P} \cap \mathcal{R}$
$\mathcal{V}_{\text{Po}}$	vertex set of Poissonized KPKVB model, i.e. Poisson process with intensity function g, resp. intensity measure μ_g
c_G	clustering coefficient of the graph G
$c_G(k)$ or $c(k;G)$	local clustering function for $k \in \mathbb{N}$ and graph G
$c_{\mathbb{H},n}(k)$	local clustering function in KPKVB
$c_\infty(k)$ or $\gamma(k)$	analytic expression for limit of clustering function of KPKVB; note that the clustering function of the infinite limit model (as an infinite graph) is undefined
c_∞ or γ	analytic expression for limit of clustering coefficient of KPKVB
$D_G(v)$ or $\deg(v)$ or $\deg_G(v)$	degree of vertex v in graph G
$N_G(k)$ or $N(k;G)$	number of degree k vertices in graph G
$N_n(k)$	number of degree k vertices in KPKVB
$N_{Po}(k)$	number of degree k vertices in poissonized KPKVB
index $\mathbb{H}, n$	refers to KPKVP graph with n vertices
upper index $*$	modified version of clustering where we divide by the expected number of vertices
$\xi = \frac{2\alpha\nu}{\pi(\alpha-\frac{1}{2})}$	
$\lambda = \frac{\alpha\nu}{\pi}$	
$\mu(y) = \xi e^{\frac{y}{2}}$	$\mu(y) = \mu(p) = \mu(\mathcal{B}_\infty(p))$
$\mu_{Po,n}(y)$	expected degree in poissonized KPKVB model of vertex with height y (in the transformed box coordinates)
$\rho(y,k) = \rho(p,k)$	$\mathbb{P}(Po(\mu(y)) = k)$
$\hat{\rho}_n(y,k) = \hat{\rho}_n(p,k)$	$\mathbb{P}(Po(\mu_n(y)) = k)$
D	degree of the typical point $(0,y)$
p_k	pmf of D, i.e. $\mathbb{P}(D=k) = \frac{2\alpha\xi^{2\alpha}\Gamma^+(k-2\alpha,\xi)}{k!}$

$P(y)$	probability that two neighbours of the typical point $(0, y)$ are adjacent, i.e. $P(y) := \mathbb{E}\mathbb{1}_{\{u_1 \in \mathcal{B}_\infty(u_2)\}}$ where u_1, u_2 are i.i.d. points in $\mathcal{B}_\infty((0, y))$ with the density $f \cdot 1_{\mathcal{B}_\infty((0,y_0))}/\mu(y_0)$.
$P(y_0, y_1, y_2)$	probability that two neighbours of the typical point $(0, y_0)$ with heights y_1 and y_2 are themselves adjacent, i.e. probability that $(0, y_0), (x_1, y_1), (x_2, y_2)$ satisfy $\lvert x_1 - x_2\rvert \leq e^{(y_1+y_2)/2}$, where x_1 and x_2 are independent uniform random variables in, respectively, $[-e^{\frac{1}{2}(y_0+y_1)}, e^{\frac{1}{2}(y_0+y_1)}]$ and $[-e^{\frac{1}{2}(y_0+y_2)}, e^{\frac{1}{2}(y_0+y_2)}]$.
a.a.s.	asymptotically almost surely.
$\xrightarrow[n\to\infty]{\mathbb{P}}$	convergence in probability
log or ln	natural logarithm with base e
$a_n \sim b_n$	$a_n = (1 + o(1))b_n$, resp. $\frac{a_n}{b_n} \to 1$
$\lvert x\rvert_m$	$\min(\lvert x\rvert, m - \lvert x\rvert)$

Curriculum Vitae

Markus Schepers was born on 16 July 1993 in Freiburg i.Br., Germany. He obtained his Abitur from Rabanus-Maurus Gymnasium in Mainz in 2012. He was awarded his bachelor's in mathematics with a minor in computer science by the University of Kaiserslautern, Germany in 2015. As part of his bachelor's, he spent an exchange semester at the National University of Singapore in 2014. He graduated with a master's degree from the University of Oxford with a distinction in 2016, with a thesis on the Kakeya problem under the supervision of Ben Green. He pursued his PhD under the supervision of Tobias Müller, spending his first year in Utrecht and moving to Groningen, The Netherlands, in 2017.

Appendix

A Meijer's G-function

Recall that $\Gamma(z)$ denotes the Gamma function. Let p, q, m, ℓ be four integers satisfying $0 \leq m \leq q$ and $0 \leq \ell \leq p$ and consider two sequences $\mathbf{a}_p = \{a_1, \ldots, a_p\}$ and $\mathbf{b}_q = \{b_1, \ldots, b_q\}$ of reals such that $a_i - b_j$ is not a positive integer for all $1 \leq i \leq p$ and $1 \leq j \leq q$ and $a_i - a_j$ is not an integer for all distinct indices $1 \leq i, j \leq p$. Then, with ι denoting the complex unit, Meijer's G-Function [39] is defined as

$$G_{p,q}^{m,\ell}\left(z \middle| \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix}\right) = \frac{1}{2\pi\iota} \int_L \frac{\prod_{j=1}^{m} \Gamma(b_j - t) \prod_{j=1}^{\ell} \Gamma(1 - a_j + t)}{\prod_{j=m+1}^{q} \Gamma(1 - b_j + t) \prod_{j=\ell+1}^{p} \Gamma(a_j - t)} z^t \, \mathrm{d}t, \tag{119}$$

where the path L is an upward oriented loop contour which separates the poles of the function $\prod_{j=1}^{m} \Gamma(b_j - t)$ from those of $\prod_{j=1}^{n} \Gamma(1 - a_j + t)$ and begins and ends at $+\infty$ or $-\infty$.

The Meijer's G-Function is of very general nature and has relation to many known special functions such as the Gamma function and the generalized hypergeometric function. For more details, such as many identities for $G_{p,q}^{m,\ell}\left(z \middle| \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix}\right)$ see [28, 38].

For our purpose we need the following identity which follows from an Mellin transform operation.

Lemma A.1. *For any $a \in \mathbb{R}$ and $\xi, s > 0$,*

$$\Gamma^{+}(-a-1, \xi/s) = G_{1,2}^{2,0}\left(\frac{\xi}{s} \middle| \begin{matrix} 1 \\ -a-1, 0 \end{matrix}\right)$$

Proof. Let $x > 0$ and $q \in \mathbb{R}$ and note that as the Γ-function is the Mellin transform of e^{-x}, by the inverse Mellin transform formula, we have $e^{-x} = \frac{1}{2\pi\iota} \int_{c-\iota\infty}^{c+\iota\infty} \Gamma(p) x^{-p} dp$ for $c > 0$ (see [20, p.196]). Applying the change of variable $p(r) = q - r$ yields $e^{-x} = \frac{1}{2\pi\iota} \int_{c+q-\iota\infty}^{c+q+\iota\infty} \Gamma(q - r) x^{r-q} dr$, then multiplying both sides with $-x^{q-1}$ gives $-x^{q-1} e^{-x} = -\frac{1}{2\pi\iota} \int_{c+q-\iota\infty}^{c+q+\iota\infty} \Gamma(q - r) x^{r-1} dr$. Now, integrating both sides gives $\int_x^\infty t^{q-1} e^{-t} dt = \frac{1}{2\pi\iota} \int_{c+q-\iota\infty}^{c+q+\iota\infty} \frac{\Gamma(q-r)}{-r} x^r dr$. On the left-hand side is the incomplete gamma function and on the right-hand side with using $-r = \frac{\Gamma(1-r)}{\Gamma(-r)}$ is the Meijer G-function, i.e. $\Gamma^+(q,x) = G^{2,0}_{1,2}\left(x \middle| \begin{matrix} 1 \\ q, 0 \end{matrix}\right)$. The claim follows by plugging in $q = -a - 1$ and $x = \frac{\xi}{s}$. □

B Incomplete beta function

Here we derive the asymptotic behavior for the function $B^-(1-z; 2\alpha, 3-4\alpha)$ as $z \to 0$, which is used to analyze the asymptotic behavior of $P(y)$, see Section 4.1.3.

Lemma B.1. *We have the following asymptotic results for* $B^-(1-z; 2\alpha, 3-4\alpha)$

1. *For* $1/2 < \alpha < 3/4$

$$\lim_{z\to 0} B^-(1-z, 2\alpha, 3-4\alpha) = B(2\alpha, 3-4\alpha).$$

2. *When* $\alpha = 3/4$,

$$\lim_{z\to 0} \frac{B^-(1-z, 2\alpha, 3-4\alpha)}{\log(z)} = -1.$$

3. *For* $\alpha > 3/4$,

$$\lim_{z\to 0} z^{4\alpha-3} B^-(1-z, 2\alpha, 3-4\alpha) = \frac{1}{4\alpha - 3}.$$

Proof. We use the hypergeometric representation of the incomplete Beta function,

$$B^-(x, a, b) = \frac{x^a}{2a} F(a, 1-b, a+1, x),$$

where F denote the hypergeometric function [49] (or see [1, Section 8.17 (ii)]). In particular we have that

$$B^-(1-z; 2\alpha, 3-4\alpha) = \frac{(1-z)^{2\alpha}}{2\alpha} F(2\alpha, 4\alpha-2, 2\alpha+1, 1-z).$$

The behavior of $F(a, b, c, 1-z)$ as $z \to 0$ depend on the real part of the sum of $c - a - b$ and whether $c = a + b$ [7] (or see [1, Section 15.4(ii)]). Since in our case a, b, c will be real it only depends on the sum of $c - a - b$. For $c - a - b > 0$ we have

$$\lim_{z\to 0} F(a, b, c, 1-z) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}, \tag{120}$$

if $c = a + b$ then

$$\lim_{z\to 0} \frac{F(a, b, c, 1-z)}{\log(z)} = -\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}, \tag{121}$$

and finally, when $c-a-b<0$

$$\lim_{z\to 0}\frac{F(a,b,c,1-z)}{z^{c-a-b}}=\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}. \tag{122}$$

In our case we have,

$$B^{-}(1-z;2\alpha,3-4\alpha)=\frac{(1-z)^{2\alpha}}{2\alpha}F(a,b,c,1-z),$$

with $a:=2\alpha$, $b:=4\alpha-2$ and $c:=2\alpha+1$. Therefore,

$$c-a-b=2\alpha+1-2\alpha-(4\alpha-2)=3-4\alpha.$$

Now if $\alpha<3/4$ then $c-a-b>0$ and hence

$$\begin{aligned}\lim_{z\to 0}B^{-}(1-z;2\alpha,3-4\alpha)&=\frac{1}{2\alpha}\frac{\Gamma(2\alpha+1)\Gamma(3-4\alpha)}{\Gamma(1)\Gamma(3-2\alpha)}\\&=\frac{\Gamma(2\alpha)\Gamma(3-4\alpha)}{\Gamma(3-2\alpha)}=B(2\alpha,3-4\alpha),\end{aligned}$$

where we used that $\Gamma(2\alpha+1)=2\alpha\Gamma(2\alpha)$.

When $\alpha=3/4$ then $c-a-b=0$ and therefore (121), together with the fact that $(1-z)^{3/2}\sim 1$ as $z\to 0$, implies that

$$\lim_{z\to 0}\frac{B^{-}(1-z;2\alpha,3-4\alpha)}{\log(z)}=-\frac{1}{2\alpha}\frac{\Gamma(6\alpha-2)}{\Gamma(2\alpha)\Gamma(4\alpha-2)}=-\frac{\Gamma(5/2)}{\frac{3}{2}\Gamma(3/2)}=-1.$$

Finally, when $\alpha>3/4$, $c-a-b=3-4\alpha<0$ and using (122) we get

$$\lim_{z\to 0}z^{4\alpha-3}B^{-}(1-z,2\alpha,3-4\alpha)=\frac{1}{2\alpha}\frac{\Gamma(2\alpha+1)\Gamma(4\alpha-3)}{\Gamma(2\alpha)\Gamma(4\alpha-2)}=\frac{\Gamma(4\alpha-3)}{\Gamma(4\alpha-2)}=\frac{1}{4\alpha-3}.$$

□

C Auxiliary approximation of a function

Lemma C.1. *For any $0<\lambda<1$ there exists a $K>0$, such that for all $0<x\leq(1-\lambda)2$*

$$\frac{1}{2}\arccos(1-x)\leq\frac{x}{\sqrt{1-(1-x)^2}}\leq\frac{1}{2}\arccos(1-x)\,(1+x)\,.$$

In particular, as $x\to 0$,

$$\frac{x}{\sqrt{1-(1-x)^2}}\sim\frac{1}{2}\arccos(1-x).$$

Proof. First we observe that for all $0 < x < 2$

$$0 < \sqrt{2x}\left(1 - \frac{x}{\sqrt{8}}\right) \leq \arccos(1-x) \leq \sqrt{2x}\left(1 + \frac{x}{\sqrt{8}}\right)$$

while for every $0 < \lambda < 1$, there exists a $K > 0$ such that for all $0 < x \leq (1-\lambda)2$,

$$0 < \frac{1}{\sqrt{2x}}\left(1 - \frac{x}{2}\right) \leq \frac{1}{\sqrt{1-(1-x)^2}} \leq \frac{1}{\sqrt{2x}}\left(1 + Kx\right).$$

It then follows that for all $0 < x \leq (1-\lambda)2$,

$$\begin{aligned}\frac{x}{\sqrt{1-(1-x^2)}} &\leq \frac{1}{2}\sqrt{2x}\left(1 + K\frac{x}{\sqrt{2}}\right) \leq \frac{1}{2}\arccos(1-x)\frac{1+Kx}{1-\frac{x}{\sqrt{8}}} \\ &\leq \frac{1}{2}\arccos(1-x)\left(1 + \frac{(K+1)x}{1-x}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{x}{\sqrt{1-(1-x^2)}} &\geq \frac{1}{2}\sqrt{2x}\left(1 - \frac{x}{2}\right) \geq \frac{1}{2}\arccos(1-x)\frac{1-\frac{x}{2}}{1+\frac{x}{\sqrt{8}}} \\ &\geq \frac{1}{2}\arccos(1-x)\left(1 - \frac{(1+\sqrt{2})x}{1+x}\right),\end{aligned}$$

which finishes the proof. □

D Some results for random variables

Let $\mathrm{Bin}(n,p)$ denote a Binomial random variable with n trials and success probability p, and $0 < \delta < 1$. Then we have the following well-known Chernoff bound.

$$\mathbb{P}\left(|\mathrm{Bin}(n,p) - np| > \delta np\right) \leq e^{-\frac{\delta^2 np}{3}}. \tag{123}$$

The following technical lemma establishes two results that are important in Section 4.7.

Lemma D.1. *Let* $\mathrm{Bin}(m,p)$ *denote a Binomial random variable with* m *trials and success probability* p*, let* $k_n \to \infty$ *be a sequence of integers such that* $k_n = o(n)$ *and fix some* $C > 0$*. Then the following holds for any* $s, t > 0$*:*

1. *for any* $0 < p_n < 1$ *such that* $|np_n - k_n| > C\sqrt{k_n \log(k_n)}$*,*

$$\mathbb{P}\left(\mathrm{Bin}(n-t, p_n) = k_n\right) = O\left(k_n^{-\frac{C^2}{3}}\right).$$

2. *for any* $0 < p_n < 1$ *such that* $|np_n - k_n| \le C\sqrt{k_n \log(k_n)}$ *and any sequence of integers* m_n *such that* $|m_n - n| \le C\sqrt{n\log(n)}$

$$\mathbb{P}\left(\mathrm{Bin}(n-t, p_n) = k_n\right) = (1 + o\left(1\right))\mathbb{P}\left(\mathrm{Bin}(m_n - s, p_n) = k_n\right).$$

Proof. First we observe that

$$\frac{\partial}{\partial x}\mathbb{P}\left(\mathrm{Bin}(m, x) = k\right) = \binom{m}{k}\left(kx^{k-1}(1-x)^{m-k} - (m-k)x^k(1-x)^{m-k-1}\right).$$

Hence, the function $x \mapsto \mathbb{P}\left(\mathrm{Bin}(m,x) = k\right)$ attains it maximum at $x = k/m$ and is strictly increasing on $(0, k/m]$ and strictly decreasing on $[k/m, 1)$.

We proceed with proving the first statement. Consider the case where $np_n < k_n - C\sqrt{k_n \log(k_n}$. Define

$$q_n = \frac{k_n - C\sqrt{k_n\log(k_n)}}{n-t}$$

and set $Y_n = \mathrm{Bin}(n-t, q_n)$. Then, since $q_n < k_n/(n-t)$ we get

$$\mathbb{P}\left(\mathrm{Bin}(n-t, p_n) = k_n\right) \le \mathbb{P}\left(Y_n = k_n\right) \le \mathbb{P}\left(Y_n > k_n - 1\right) = \mathbb{P}\left(Y_n > (1+\delta_n)(n-t)q_n\right),$$

with

$$\delta_n = \frac{k_n - 1 - nq_n}{(n-t)q_n} = \frac{C\sqrt{k_n\log(k_n)} - 1}{k_n - C\sqrt{k_n\log(k_n)}}.$$

By a Chernoff bound we get

$$\mathbb{P}\left(\mathrm{Bin}(n-t, p_n) = k_n\right) \le \mathbb{P}\left(Y_n > (1+\delta_n)nq_n\right) \le e^{-\frac{\delta_n^2 nq_n}{2}} \le e^{-\frac{\delta_n^2 nq_n}{3}}.$$

The result now follows by observing that $\delta_n^2 nq_n = \Omega\left(C^2\log(k_n)\right)$.

For the case $np_n > k_n - C\sqrt{k_n\log(k_n}$ we set $q_n = (k_n + C\sqrt{k_n\log(k_n)}/n > k_n/n$ and

$$\delta_n = \frac{nq_n - (k_n + 1)}{nq_n} = \frac{C\sqrt{k_n\log(k_n)} - 1}{k_n + C\sqrt{k_n\log(k_n)}}.$$

Then, $p_n > q_n > k_n/n$ and $\delta_n^2 nq_n = \Omega\left(C^2\log(k_n)\right)$. Thus by another Chernoff bound

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}(n-t, p_n) = k_n\right) &\le \mathbb{P}\left(Y_n = k_n\right) \le \mathbb{P}\left(Y_n < k_n + 1\right)\\ &= \mathbb{P}\left(Y_n < (1-\delta_n)nq_n\right) \le e^{-\frac{\delta_n^2 nq_n}{3}} = O\left(k_n^{-\frac{C^2}{3}}\right).\end{aligned}$$

Now let us prove the second statement. For this we first note that by Stirling's formula

$$\binom{n-t}{k_n} \sim (2\pi k_n)^{-1/2}\left(\frac{k_n}{n-t}\right)^{-k_n}\left(1 - \frac{k_n}{n-t}\right)^{k_n-(n-t)},$$

and similarly for $\binom{m_n-s}{k_n}$. Therefore

$$\begin{aligned}
&\binom{n-t}{k_n}\binom{m_n-s}{k_n}^{-1} \\
&= (1+o(1))\left(\frac{n-t}{m_n-s}\right)^{k_n}\left(1-\frac{k_n}{n-t}\right)^{k_n-(n-t)}\left(1-\frac{k_n}{m_n-s}\right)^{m_n-s-k_n}
\end{aligned}$$

and thus

$$\begin{aligned}
&\frac{\mathbb{P}\left(\mathrm{Bin}(n-t,p_n)=k_n\right)}{\mathbb{P}\left(\mathrm{Bin}(m_n-s,p_n)=k_n\right)} \\
&= (1+o(1))\left(\frac{n-t-k_n}{n-t-p_n}\right)^{k_n-(n-t)}\left(\frac{m_n-s-k_n}{m_n-s-p_n}\right)^{k_n-(m_n-s)}
\end{aligned}$$

Let us rewrite the first multiplicative term as

$$\left(\frac{n-t-k_n}{n-t-p_n}\right)^{k_n-(n-t)} = \left(1-\frac{k_n-p_n}{n-t-p_n}\right)^{k_n-(n-t)} := (1-x_n)^{k_n-(n-t)}.$$

Then, using that for all $-1/2 \le x \le 1/2$, $-x-x^2 \le \log(1-x) \le -x$, we get

$$e^{-(k_n-(n-t))\left(x_n+x_n^2\right)} \le \left(\frac{n-t-k_n}{n-t-p_n}\right)^{k_n-(n-t)} \le e^{-(k_n-(n-t))x_n}.$$

Similarly,

$$e^{-(m_n-s-k_n)\left(y_n+y_n^2\right)} \le \left(\frac{m_n-s-k_n}{m_n-s-p_n}\right)^{m_n-s-k_n} \le e^{-(m_n-s-k_n)y_n},$$

where

$$y_n := \frac{k_n-p_n}{m_n-s-p_n}.$$

We first show that

$$\lim_{n\to\infty} e^{-(k_n-(n-t))x_n}e^{-(m_n-s-k_n)y_n} = 1. \tag{124}$$

With some algebra we get

$$\begin{aligned}
&-(k_n-(n-t))x_n-(m_n-s-k_n)y_n \\
&\quad= \left(1-\frac{k_n-p_n}{n-t-p_n}\right)(k_n-p_n) - \left(1-\frac{k_n-p_n}{m_n-s-p_n}\right)(k_n-p_n) \\
&\quad= (k_n-p_n)^2\left(\frac{1}{m_n-s-p_n}-\frac{1}{n-t-p_n}\right)
\end{aligned}$$

$$= (k_n - p_n)^2 \frac{n - m_n - t + s}{(m_n - s - p_n)(n - t - p_n)}.$$

Now by our assumptions

$$(k_n - p_n)^2 = \Theta\left(k_n \log(k_n)\right),$$

while

$$m_n - s - p_n = \Theta\left(n - t - p_n\right) = \Theta\left(n\right)$$

and

$$-C\sqrt{n\log(n)} \leq n - m_n \leq C\sqrt{n\log(n)}.$$

Therefore we conclude that

$$(k_n - p_n)^2 \frac{n - m_n - t + s}{(m_n - s - p_n)(n - t - p_n)} = \Theta\left(\frac{k_n \log(k_n)\sqrt{n\log(n)}}{n^2}\right)$$

from which (124) follows.

In a similar way it follows that

$$\lim_{n\to\infty} e^{-(k_n-(n-t))x_n^2} e^{-(m_n-s-k_n)y_n^2} = 1,$$

which implies that

$$\lim_{n\to\infty} \frac{\mathbb{P}\left(\mathrm{Bin}(n-t, p_n) = k_n\right)}{\mathbb{P}\left(\mathrm{Bin}(m_n - s, p_n) = k_n\right)} = 1,$$

as thus finishes the proof. □

E Code for the simulations

The simulations of the clustering coefficient and function in the KPKVB model were done using Wolfram Mathematica 11.1. The simulation dots for the clustering coefficient in Figure 1.5 were generated by the following code (where in the second line, the entire script was also run for the values `nu=1` and `nu=0.5`):

```
1  n=10000;
2  nu=2;
3  R=2*Log[n/nu];
4  plotpoints=20;
5  reps=100;
6  Plotingdataalpha = ConstantArray[0,{plotpoints,2}];
7  SeedRandom[1];
8  For[z=1,z<=plotpoints,z++,a=0.4+z (4.6/plotpoints); sum=0;
9      For[r=1,r<=reps,r++,V = ConstantArray[0,{n,2}];
10         For[i=1,i<=n,i++,
```

```
                V[[i,1]]=RandomReal[{-Pi ,Pi }];
                V[[i,2]]=ArcCosh[RandomReal[{0,1}](Cosh[a*R]-1)
                    +1]/a];
            A= ConstantArray[0,{n,n}];
            For[i=1,i<=n,i++,
                For[j=1,j<=n,j++,
                    If[Cosh[V[[i,2]]]Cosh[V[[j,2]]] - Sinh[V[[i
                        ,2]]] Sinh[V[[j,2]]] Cos[Abs[V[[i,1]] -V[[
                        j,1]]]] <= Cosh[R] && i != j,A[[i,j
                        ]]=1,A[[i,j]]=0]]];
            g = AdjacencyGraph[A];
            sum=sum+MeanClusteringCoefficient[g]];
        Plotingdataalpha[[z,1]]=a;
        Plotingdataalpha[[z,2]]=1.0*sum/reps;]
Print[Plotingdataalpha]
```

The simulation dots for the clustering function in Figure 1.8 were generated by the following code (where in the third line, the entire script was also run for the values `nu=1` and `nu=0.5`):

```
n=10000;
a=0.8;
nu=2;
R=2*Log[n/nu];
plotpoints=24;
reps=100;
Plotingdatak = ConstantArray[0,{reps,plotpoints,2}];
SeedRandom[1];
For[r=1,r<=reps,r++,V = ConstantArray[0,{n,2}];
    For[i=1,i<=n,i++,
        V[[i,1]]=RandomReal[{-Pi ,Pi }];
        V[[i,2]]=ArcCosh[RandomReal[{0,1}](Cosh[a*R]-1)+1]/
            a];
    A= ConstantArray[0,{n,n}];
    For[i=1,i<=n,i++,
        For[j=1,j<=n,j++,
            If[Cosh[V[[i,2]]]Cosh[V[[j,2]]] - Sinh[V[[i,2]]]
                Sinh[V[[j,2]]]Cos[Abs[V[[i,1]] -V[[j,1]]]]
                <= Cosh[R] && i != j,A[[i,j]]=1,A[[i,j
                ]]=0]]];
    g = AdjacencyGraph[A];
    For[k=1,k<=plotpoints,k++,
        sum=0;
        result=0;
```

```
        nrdegk=0;
        For[v =1,v<=n,v++;
            If[VertexDegree[g,v]==k+1,
                result=result+LocalClusteringCoefficient[g,v];nrdegk++]];
        Plotingdatak[[r,k,1]]=k+1;
        If[nrdegk>0,Plotingdatak[[r,k,2]]=1.0*result/nrdegk];]
Print[Mean[Plotingdatak]];
```